Toward High Throughput Screening of Assembly-Optimised Ferritin Nanocage Libraries Applied to Nanotechnology

Cornell, Thomas Andrew

Awarding institution:
King's College London

Toward High Throughput Screening of Assembly-Optimised Ferritin Nanocage Libraries Applied to Nanotechnology

Thomas Andrew Cornell

King's College London

Department of Chemistry

School of Natural and Mathematical Sciences

A thesis submitted to King's College London
in partial fulfilment of the requirement for the degree of
Doctor of Philosophy

2014

Blessed is the Second mouse for he shall inherit the Cheese.

Anon.

Acknowledgements

First and foremost I must acknowledge my supervisor, Dr. *Brendan Patrick Orner*. His patience, teachings and wisdom were perfectly tailored throughout my PhD, always leaving me engrossed in the science's we tried to understand and eager to learn more. I have appreciated the opportunities and freedoms he has granted to explore and develop my own ideas, and his commitment as a mentor to ensure that I always achieved my best. I am also truly grateful to him for not only helping me to write this piece of work, but to of guided me through it from its inception.

I would like to thank my second supervisor Prof. *Brian Sutton* for his mentorship and advice. My work could not have proceeded without the kind discussions and insights from Dr. *Anh Tuan Phan*, Dr. *Ralf Jauch*, Dr. *Ali Miserez*, Prof. *Roger Morris*, Dr. *Sarah Barry,* Prof. *Phil Blower,* Dr. *Maggie Cooper* and Dr. *Bobbi Fleiss*. I thank the core facilities; support and teaching laboratory staff at King's College London and Nanyang Technological University for their help. My appreciation goes to the KCL chemistry department start up grant and the NTU SINGA scholarship award for funding me during my PhD. My sincere thanks go to my fellow graduate students who have helped me along this journey, Dr. *Rongli Fan*, Dr. *Zhang Yu*, *Jon Palmer* and *Cathy Hubert*. My heartfelt gratitude goes to Dr. *Maziar Ardejani* and *Chris Chamberlain* who both taught me so much, my PhD would have been very different without your support. I appreciate the assistance of my undergraduate mentees- *Joy Sebastian*, *Guo Feng* and *Sara Rodriguez* for their support in making nanoparticles. *Stephanie Newland*, *Hugh McGirr* and *Alice Rerat* who helped get FlAsH off the ground, *Abbas Bacha* for his dedication to the development of an iron assay and to everyone else I have had the great pleasure of working with. My highest appreciation goes to *Fu Jing*, who wind, rain or shine was in the laboratory helping me. Your friendship and enthusiasm are an inspiration.

Last but not least, I wish to thank those most important to me, my parents, *Maggie* and *Simon Cornell*, and my family for their unending love, compassion and support. I hope I

have made you proud; I could not have aimed so high without you. Lastly I wish to thank my partner *Kenji Teo,* whose support has never faded even while being 7000 miles apart. I appreciate your patience, love and understanding, even with my insistence that fluorescence data makes for good dinner conservation.

Table of contents

Toward High Throughput Screening of Assembly-Optimised Ferritin Nanocage Libraries Applied to Nanotechnology

Chapter 2

Chapter 5

Direct detection of mini-ferritin assembly with the fluorescent probe FlAsH-EDT$_2$ 144

Chapter 6

Bipartite tetracysteine display for high throughput screening of Dps libraries in living bacteria.....181

List of figures and tables

Abstract

Protein nanocages have been directed toward a myriad of applications such as drug delivery and material science as well as model systems to study protein folding and self-assembly. These pursuits, because they have conventionally been explored through often indirect techniques which typically are implemented in a protracted and high-resource intensive manner, have been limited by the inability to design rapidly novel and unique properties of these proteins or to discover conditions most ideal for their efficacy. This thesis describes experiments performed to explore protein nanocage applications and fundamental self-assembly in a traditional sense while providing a path for the broadening of these goals by developing a novel technique to explore the properties and conditions of nanocage assembly quickly and directly.

Chapter 1 provides an introduction to protein quaternary structure and nanocage assembly with a focus on the ferritin family of proteins. Chapter 2 describes the attempt to apply protein nanocages to nanotechnology. Chapter 3 reports the application of x-ray crystallography to understand the role of sub-tertiary structure in the generation of protein nanocage quaternary structure. However, although this work advanced nanoscience and provided fundamental insight, it became evident that these strategies could be augmented through the development of a more rapid screening technique for protein cage assembly. Thus, Chapter 4 and Chapter 5 develop this technique with two different protein nanocages *in vitro*, while Chapter 6 further expands it to allow the high throughput screening of protein libraries *in vivo*. This method could help to expand and hasten our understanding of protein quaternary structure as well as to facilitate the development or enhancement of nanocage applications where tailored properties would be desired for the generation of size selective nanoparticles or for the transport and delivery of drugs.

Chapter 1

Introduction

1.1 Introduction

The generation of complex structures through the assembly of simpler and smaller building blocks is ubiquitous to Nature. The resulting structures are determined by both the inherent properties of the building blocks and the assembly process itself[1]. The interactions between the subunits and restrictions imposed by requirements from the end-structure as it forms, can also come into play. The research projects documented in this thesis, while diverse, are conceptually united in that they arise from the desire to understand or to exploit the assembly of complex structure.

An analogy to the assembly of complex structures is a comparison to a brick house. The size, shape and composition of the individual bricks can determine the resulting structure of the building. Understanding the fundamental properties of the bricks can lead to innovations in architecture, while new innovations can also be achieved by inventing bricks with enhanced properties. A house, while a hollow cubic arrangement of bricks, can also be a home which provides shelter and a unique environment from that experienced by the outside of the house.

While the analogy to a brick home is rather simplistic, it does exemplify the assembly of complexity from many smaller, less complex, building blocks. Examples from biology are much more complicated and less able to be completely understood and described. Multi-cellular organisms, for example, and some single celled organisms under certain circumstances like with biofilms,[2] are assembled from cells which in turn are assembled from systems generated by the interactions of bio-molecules. Although each of these levels of building blocks (organs and tissues, cellular niches, cells, metabolic and genetic systems, quaternary structure, molecules) are complex in their own right, only in combination can certain functions be realised. Often, the function of the collective is unexpected or non-obvious even by close understanding of the individual building blocks and only through their interaction, can the new properties of the complex structure emerge.

Interestingly, a similar emergence of new properties is the focus of research into nanomaterials[3]. Although chemistry has matured to the extent that it is possible to predict molecular properties, it is nearly impossible to anticipate the bulk properties when these molecules are assembled into materials. It is in the nanoscale, when a threshold of atoms or molecules are assembled, where bulk properties of a material start to emerge. Thus, nanoscaled materials are extremely sensitive to size and shape which both are related to the number of building blocks and how they are assembled.

With the rise of nanotechnology, we are beginning to investigate and explore the complex assembly of building blocks however, in order to better exploit assembled complex systems, more fundamental understanding of the building blocks themselves and their process of assembly is required.

1.2. Organised assembly through protein-protein interactions

Constructing complex structures from the hierarchical assembly of simpler building blocks is fundamental to protein function. Protein assembly (quaternary structure) is complicated in that it is often coupled to protein folding (tertiary structure)[4, 5]. Although some proteins require the assistance of chaperones to assist in assembly[6], most quaternary structure can form spontaneously, underlining the fact that the information required for assembly is stored in the folded polypeptide sequence and the resulting projection of surface functionality. The ability to engineer amino acid polypeptides with programmed folding and functional assembly has been explored[7] with some limited success (see section 1.8).

Quaternary structure formation is fundamental to almost all cellular processes. The interactions between proteins has been shown to be important for cell signalling[8], often further implicated in cancer[9], and for essential functions such as DNA replication[10]. Quaternary structure also provides the creation of large protein assemblies. For example, one of the first observed proteins, actin[11], can form long filaments from a single subunit, which shows rapid

assembly dynamics,[12] and controls cell shape and movement. Similarly the activity of tubulin in forming flagella, regulating facilitating cytoskeleton dependent cell division is dependent on its assembly[13].

While many protein assembly processes are fundamental to life, they can also be detrimental. Amyloids fold into cross β-strands that assemble into fibrils that eventually precipitate as insoluble plaques[14]. These large protein aggregates are the causes of several human diseases. For example, amyloid precursor protein, which is proteolysed into β-amyloid, can aggregate in the brain and is believed to lead to Alzheimer's disease[15]. In addition, prion proteins can also misfold and aggregate. Interestingly, these aggregates appear to catalyse the aggregation of non-aggregated prion protein with such mechanisms being involved in diseases such as Creutzfeldt–Jakob disease (CJD)[16].

Nature's strategy to assemble large, complex protein structures often takes advantage of symmetry[1]. The synthetically least intensive way to generate complexity is to assemble simple building blocks into structures with high symmetry. Repeating subunits can be utilised and types of interactions between them can be minimised so that complex solutions to problems can be realised with a higher than expected evolutionary efficiency. Furthermore, these symmetric interactions are often focused at 'hot spot' areas and not spread along the length of the interaction[17]. This further complicates their characterisation and eventual design as it can often remain unclear where these areas are located without extensive investigations.

1.3. Protein nanocages

One of the most common families of highly symmetric assembled protein structures are the nanocages where monomers are assembled into symmetrical capsules with large, hollow central cavities. One type of protein nanocage are virus capsids. Virus assemblies range in size from the small protein capsids of Porcine circovirus (17 nm diameter)[18], to the giant virus, *Pithovirus* (1.5 μm)[19]. In general, virus capsid structure is either spherical with

icosahedral symmetry, such as in cowpea chlorotic mottle virus[20] and the adenoviruses[21], or non-spherical with helix-type structures such as in tobacco etch virus[22].

Viruses are not the only examples of protein nanocages. Among these is the ferritin superfamily (see section 1.4) which help to buffer cellular iron. Heat shock proteins, which form large nanocages of a large variety of sizes (from 8 – 80 kDa)[23], that not only protect the cell from stress, but also act as chaperones. Lumazine synthase assembles from 180 monomers into an icosahedral protein nanocage and is used by *Aquifex aeolicus* in the formation of riboflavin[24]. Another interesting nanocage, whose function is still poorly understood, is the vault protein. Measuring 34 nm across and 60 nm long, it is one of the largest non-viral protein complexes[25, 26] (see Figure 1.1 for examples of protein nanocages).

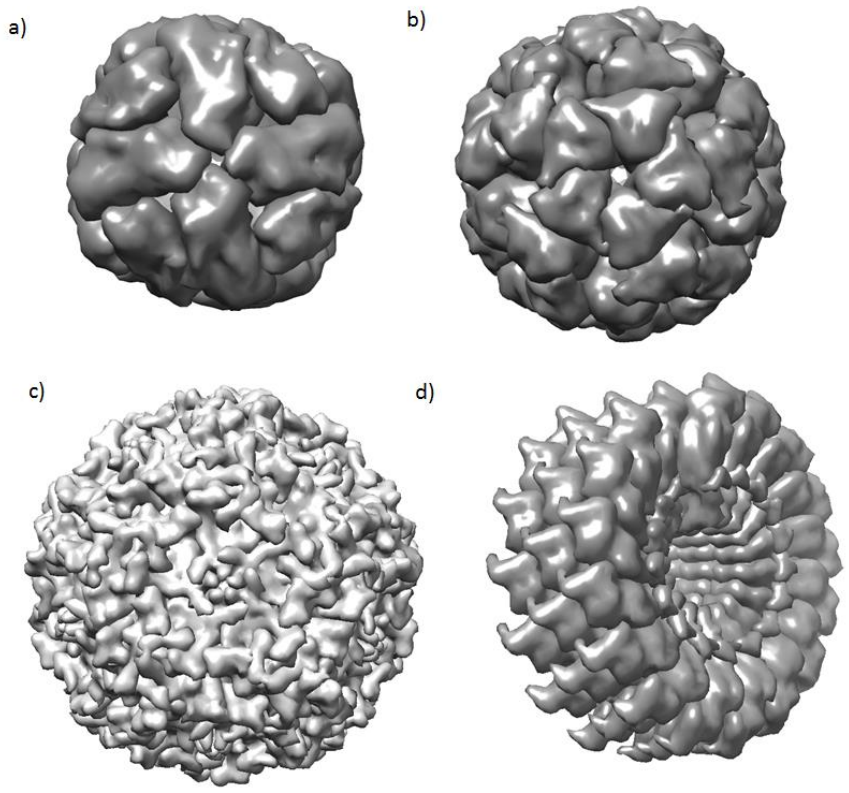


Figure 1.1 – Structures of protein nanocages and quaternary related structures. (Not to scale.) (a) Ferritin 24-mer (PDB:1BFR[27]), (b) Lumazine synthase 60-mer (PDB:1HQK[28]), (c) B19 parvovirus capsid (PDB:1S58[29]), (d) Tobacco mosaic virus (PDB:3VTM[30]). Images made using Chimera[31].

Many applications exploiting protein nanocages rely on the fact that they have three distinctively different surfaces, the surfaces on the outside of the nanocage, on the inside of the hollow cavity, and the surfaces making up the protein-protein interfaces, that can be engineered and utilised to generate constructs with multiple functions[7, 32] (Figure 1.2). The protein-protein interfacial surfaces can be modified to change the assembly properties and possibly even the overall structure of the assembled nanocage itself. The outer protein nanocage surface has been chemically or genetically modified to control cellular localisation[33] and stability. It has even been functionalised so that the protein nanocage can be applied to novel applications like for MRI contrast agents[34]. The inner cavity surface is usually optimised for the functions of the native nanocage, such as metal mineralisation or DNA delivery, but can be augmented for new functions. The ability to target and manipulate either one of all of these very different surfaces provides a high level of control for the expansion of nanocages for novel applications.

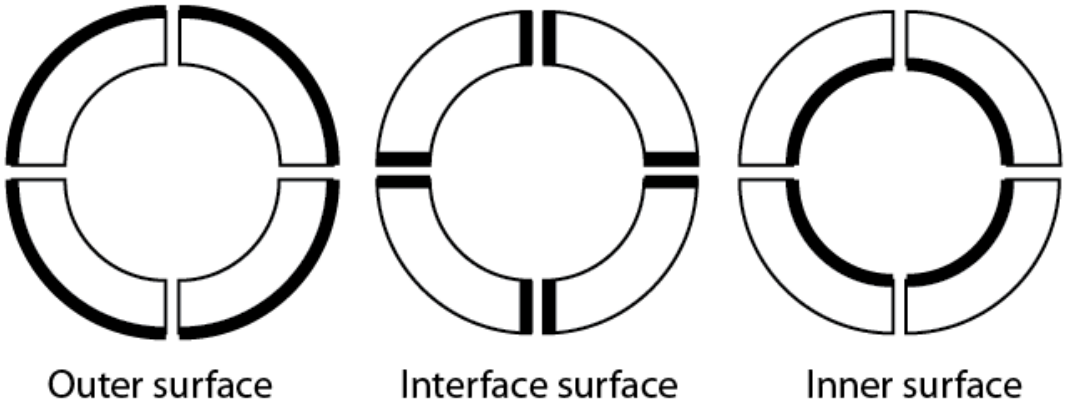


Figure 1.2 – Schematic representation highlighting the three distinctly different surfaces that can be altered in protein nanocages. (left) the outer surface, (middle) the surface located between protein interfaces and (right) the inner surface[35].

1.4. Ferritins

Ferritin proteins are well characterised and have been the subject of much nanocage research. They are also the focus of this thesis. The ferritin proteins form well-defined nanocages through symmetrically related interactions. Ferritins are evolutionarily ubiquitous,

seen in all forms of life and are employed by the cell to help control cellular iron homeostasis[36]. They generally are selectively permeable to iron ions which are mineralised and stored in the ferritin inner cavity. Ferritins have been studies for many years, and the first crystal structure of a ferritin was obtained in 1975[37]. Since then many more ferritin structures have been solved and their regulation and function have been well documented. Interestingly, although the ferritins share little sequence similarity, their tertiary structures are strikingly homologous[38]. This structural similarity is carried through to the quaternary structure with only a few distinct subfamilies.

1.4.1. Maxi-ferritins

Maxi-ferritins are assembled from twenty-four monomer subunits and form a nanocage with octahedral symmetry[39] (Figure 1.3). The most commonly observed maxi-ferritin assembly intermediate, is a two-fold symmetric anti-parallel dimer. It is believed that nanocage formation occurs from the assembly of these subunits around three- and four-fold axis of symmetry[40, 41]. Maxi-ferritins assemble into a nanocage of octahedral symmetry with an exterior diameter normally around 12 nm and an interior diameter of 7 nm[27]. An exception to this rule is the ferritin from archaeon, *Archaeoglobus fulgidus* which assemblies into a 24-meric nanocage but with tetrahedral symmetry[42].

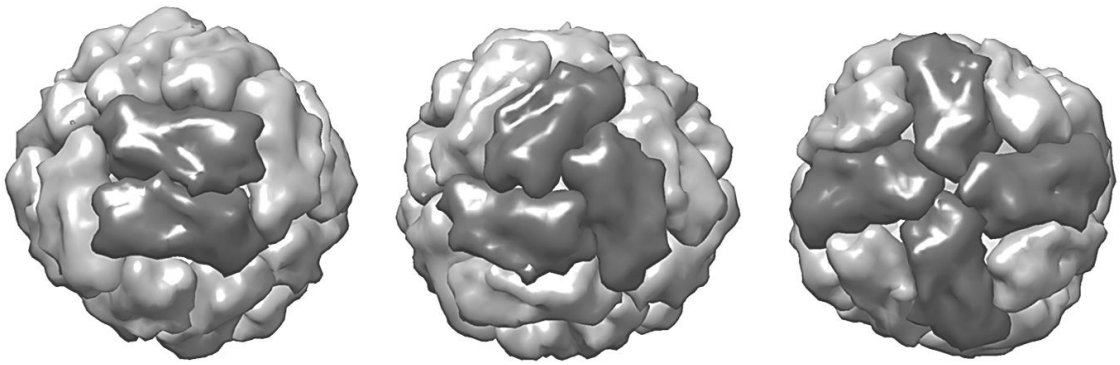


Figure 1.3 – The structure of a typical maxi-ferritin, *E.coli* Bfr (PDB:1BFR[27]) highlighting (in dark grey) interactions around its symmetry axes. (left) two-fold symmetry axis, (middle) the three-fold symmetry axis and (right) the four fold symmetry axis which is unique to maxi-ferritins. The dimer highlighted at the two-fold axis is thought to be the major assembly intermediate.

Maxi-ferritins can be organised into two subgroups, the ferritins (Ftn) and the bacterioferritins (Bfr). While ferritins are seen in all three kingdoms of life, bacterioferritins only exist bacteria and fungi[43, 44]. Interestingly, the mammalian ferritins natively assemble into heterooligomers and are formed by the combination of two heavy and light chain monomers[45]. Although, homooligomers can also assemble *in vitro*. Bfr, on the other hand, only forms homooligomers. The most striking difference between these two subgroups is that Bfr contains twelve heme groups each bound in the middle of the dimer subunit at the protein-protein interface[44].

Like all nanocages, an important feature of ferritins is that they have three surfaces available for functionalisation and engineering. The inner and outer surfaces have been examined and altered to perform new functions as discussed in section 1.5. The interfacial surfaces have been less studied. *E. coli* Bfr exists in solution as a combination between the nanocage and the dimer oligomerisation states. Our laboratory has investigated the role these surfaces play in controlling the distribution between these two species[41, 46, 47]. Through this work, we have developed proteins, with mutated protein-protein interfaces, that favour one oligomerisation state over the other. These proteins, will serve as the basis for control experiments employed in this thesis (Chapter 4).

1.4.2 Mini-ferritins

A second ferritin subfamily is made up by the mini-ferritins. One type of mini-ferritin is the DNA binding protein from starved cells (Dps). This protein shows a similar tertiary structure to the other ferritin, but it differs considerably in its quaternary structure. Dps nanocages are made up of twelve monomers. These nanocages have tetrahedral symmetry and exterior and interior diameter of 9 nm and 4.5 nm respectively[48]. Similar to maxi-ferritins, Dps is thought to assemble through dimer two-fold symmetric intermediates. The assembled nanocage of Dps also presents two types of three fold symmetry (“Dps-like” and “ferritin-like”)

(Figure 1.4). In addition to the storage of iron, Dps has been shown to bind DNA, and this binding plays a role in it providing protection from cellular oxidative stress[49]. In addition, the Orner laboratory has also generated protein-protein interfacial mutants of the Dps nanocage that can affect assembly, which will be used as controls in this thesis[50] (Chapter 6).

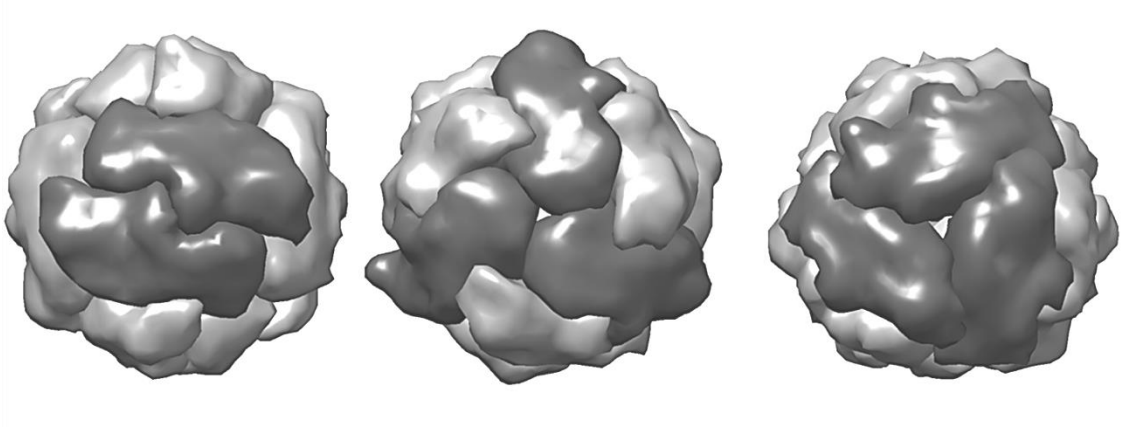


Figure 1.4 - The structure of a typical mini-ferritin, *E.coli* Dps(PDB:1DPS[51]) highlighting (in dark grey) interactions around its symmetry axes. (left) two-fold symmetry axis (dimer), (middle) the Dps-like three-fold symmetry axis and (right) the ferritin-like three-fold symmetry axis. The dimer highlighted at the two-fold axis is thought to be the major assembly intermediate.

1.4.3. Ferritins and their role in iron mineralisation and storage

The control of cellular iron homeostasis is complex. First, the membrane receptor transferrin binds iron and internalises it to the cell[52]. The soluble Fe^{2+} is then released from transferrin and stored in the ferritins. Ferritins oxidize Fe^{2+} to Fe^{3+} with H_2O_2 and store the iron as a ferrihydrite mineral inside the protein nanocage cavity[53]. Maxi-ferritins can in general hold up to 4500 iron atoms while mini-ferritins can only store 1500[54]. Each subgroup also differs in how iron is mineralised. Within the Ftn subgroup, only the heavy chain monomers are catalytic, and the light chain monomers only play a structural role in assembly. However, each Bfr monomer has ferroxidase activity, but for Dps, dimerization is required to form an intact active site[54].

1.4.4. Ferritins and their roles in disease

Ferritins can be recognised and taken up into cells from serum through specific cell surface receptors[55]. This function has led to the belief that this nanocage could be used as a delivery agent for therapeutics, a topic which is further explored in Chapter 2[56]. However, the process by which ferritin localises to serum is still poorly understood and could potentially play an important role in disease. Neuroferritinopathy, a condition tied to the aggregation of ferritin monomers into large fibres follows a marked increase in iron concentration throughout the nervous system[57, 58] and patients present symptoms similar to Alzheimer's[59]. The aggregates are favoured by nanocage destabilising mutations. This ability of ferritins to aggregate, could be a challenge to any attempt to use these nanocages for various applications (see Chapter 2).

1.5 Nanocage functionalisation

Protein nanocages have been the focus of much recent research and characterisation which is leading to their use in novel applications. Nanocage proteins, because of their defined size and shape, have been used as size-constrained reactors to afford nanoparticles with very narrow polydispersities. In addition, they provide the resulting encapsulated particles with enhanced solubility, protect the particles from aggregating, and enable the further manipulation of the particles through bioconjugate and fusion chemistries[60, 61]. To date, gold[62, 63, 64, 62, 65,] silver[66], copper[67], cadmium[68], palladium[69] and platinum[70] nanoparticles have been formed inside of ferritins, a topic which is further explored in Chapter 2. Gold nanoparticles in particular have now been shown to be useful as catalysts, in contrast enhancing optical sensors and for medical diagnosis[71]. Iron oxide nanoparticles have been formed inside the protein nanocage lumazine synthase[72]. Moreover, the formation of a tungsten core inside cowpea chlorotic mottle virus[73] has also been achieved. While other nanocages have been used for nanoparticle formation, the well characterised structure and native mineralisation activity make ferritins ideal for similar applications.

Nanocages are beginning to find other applications as well. For example they have been used to sequester toxic cargo to enhance cell survival[74] have been part of contrast enhancing agents for MRI[34, 75]. Interest in protein nanocages as drug delivery agents is also developing[76]. However applied research such as this is often hampered by a lack of fundamental understanding of the chemical and physical properties of these proteins which are mainly dependent on the assembly process and the protein-protein interactions that strongly influence it. However most of the techniques to characterise these properties are either time-intensive, indirect, or not amenable to discovering protein nanocages with new properties via high throughput screening. Because the majority of this thesis (Chapters 4, 5, and 6) focuses on the development of a rapid and direct assay for protein nanocage assembly, we will discuss the methods that have been developed to detect protein-protein interactions in the next section.

1.6. Methods for the detection of quaternary structure

Protein-protein interactions are fundamental to multiple cellular processes[8] and are playing a larger role in materials and bio-medical applications[77]. Thus, several techniques have been developed to detect and characterise interactions between proteins. These techniques have helped to establish the fundamental and disease-based roles of these interactions and where in the cell they take place. Several families of techniques have been established, some of which have single molecule resolution.

1.6.1. Detection of protein-protein interactions with classical protein characterisation techniques

Techniques that were developed to characterise proteins can likewise be used to detect when that character changes, such as when two or more proteins assemble and interact. When proteins interact their overall size tends to increase and many methods can detect that change. Methods such as size exclusion chromatography (SEC) and native

polyacrylamide gel electrophoresis (native PAGE) which separate proteins based on their size and dynamic light scattering (DLS) which can determine the size of large particles based on the light they scatter have been used to detect the change in size of a complex when protein-protein interactions occur. These three techniques have been used extensively with nanocages, and we also employ them in the research described in this thesis[40, 41, 46, 47]. While these methods are very useful for monitoring quaternary structure, they almost always require purified proteins and are limited to low throughput *in vitro* analysis.

Techniques that characterise the thermodynamics of proteins can also be used to evaluate quaternary structure. Isothermal titration calorimetry (ITC) measures the release of energy during the exothermic interaction between two binding partners[78]. While most notably used for drug discovery, ITC can monitor the interactions between biomolecules via the titration of one to the other[79]. This technique can not only help calculate the binding affinity of the interacting pair but also their stoichiometry and thermodynamic parameters describing their association. Differential scanning calorimetry (DSC), monitors the change in specific heat capacity of a solution which can be characteristic if proteins are interacting[80]. Although not widely adopted, it has been used to monitor the interaction between ribonuclease A (RNase) S protein and the RNaseA S peptide[81]. Our laboratory has used DSC to characterise ferritin protein nanocages. Although useful, ITC and DSC like the above techniques, require purified proteins, often in large quantities, and are limited to low throughput *in vitro* screening.

Another technique that is more frequently used for protein characterisation but can be modified to examine protein-protein interactions is mass spectrometry (MS)[82]. This can be challenging because large proteins are not only large in mass but they are also can be unstable in the gas phase and under ionization conditions[83]. This technique has been used to map the Wnt signalling pathway[84] as well as assessing protein interactions in cancer cells using affinity-purification mass spectrometry (AP-MS)[84]. MS techniques generally are rapid and use only a

small amount of sample, and advances in proteomics have advanced MS to high throughput. However, these proteomics strategies often involve enrichment steps that have not yet been developed for protein nanocages[85]. With that said, well characterised or purified protein samples could benefit from mass spectrometry analysis.

Techniques employing surface plasmon resonance (SPR) detection have gained much attention due to the fact that kinetic binding information can often be obtained[86]. SPR works by passing a beam of light through a prism at a resonance angle toward a thin layer of gold or silver. Some of this light interacts with the electron cloud of the metal surface generating a plasmon. This plasmon creates a small electric field on both sides of the thin layer and any change in environment near this field, such as binding of a protein to an immobilised target, creates a measureable change in resonance angle dependent upon a change in refractive index. From SPR, the association and dissociation constants and therefore the binding constant can be obtained, as well as the stoichiometry of the interacting parties. SPR has been used to monitor the concentration of serum ferritins, through its interactions with an antibody[87] as well as monitoring the interactions of p53 and p73[88]. However this method is still limited to relatively low throughput *in vitro* conditions, requiring purified proteins and it is unclear how it could be applied to the assembly of protein nanostructure.

While all of the above methods can help characterise protein-protein interactions they are all low throughput, often requiring large quantities of purified protein (with the exception of MS), and are limited to *in vitro* analysis. However, there are several techniques that have been focused more directly toward observing quaternary structure.

1.6.2. Anti-body based techniques (ELISA and Immunoprecipitation)

Antibodies can be raised against a large variety of biological and chemical targets and can be used to tag many molecules and structures inside cells[89]. This ability to be able to target almost anything makes antibodies incredible tools for the analysis of proteins. Early examples

of antibody targeting include Western blots developed by Towbin and co-workers which used antibodies to tag selective protein targets after their transfer to a nitrocellulose film[90]. This method was expanded upon to help purify proteins from cellular lysate samples. Immunoprecipitation requires the use of an antibody that has been raised against a particular protein target. Upon combining this antibody with a protein mixture, the antibody molecules can swarm the protein target, causing the target protein to precipitate from the solution due to the large molecular weight of the complex. Alternatively, the antibody can be fused to a purification tag and the target protein can be purified through a 'pull down' experiment. Quaternary structure can be analysed through co-immunoprecipitation. Similar to using affinity-purification mass spectrometry (see above), any protein that interacts to the target protein is co-purified with the target. Co-immunoprecipitation has been used to monitor the oligomersation of G protein coupled receptors[91] and to map protein interaction networks in yeast[92].

In 1971, Perlman and co-workers published the first reported application of the enzyme-linked immunosorbent assay (ELISA) which linked several existing antibody technologies and many variations of the ELISA have since been developed[93]. One version of ELISA works by first immobilising a target protein or ligand to the surface of a well. Different wells are then washed with potential binding partners which all present an antibody binding epitope. Any protein that binds to the original target remains immobilised while all unsuccessful binders are washed away. An antibody is added that binds to the antibody binding epitope. A secondary antibody that was raised against the primary antibody is then added. This secondary antibody is fused to an enzyme, most commonly a peroxidase, which allows for the turnover of chromogenic substrate leading to a coloured observable response. Therefore, any protein that binds to the target can be identified through this colourmetric assay. Importantly the turnover of the enzyme allows an amplification of a signal from a binding event. ELISA has been modified over its lifetime, for example to use just a single

antibody-peroxidase, or a fluorescent probe fusion. Although ELISA often requires extended preparation times, the commercial availability of pre made plates has increased their utility[94]. ELISA's have been extensively used to screen for the presence of infectious agents. A HIV assay relies on the detection of the HIV protein capsid using an antibody immobilised to a well[95].

This technique provides advantages over several of the ones discussed above, as the throughput of ELISA is much larger. However, both this technique and immunoprecipitation can only be performed *in vitro*, although they do work in complex solutions like lysates. The requirement of large amounts of antibodies can cause them to be prohibitively expensive. In addition, they often also need extended and repetitive washing steps which all reduce the ability of these methods to be expanded to screen a large number of proteins quickly in the absence of robotics.

1.6.3. Fluorescence Polarisation

Fluorescence polarisation (FP)[96] is a commonly used *in vitro* technique. In this method, polarised light is use to excite a molecular probe attached to one protein of a complex. The degree that the polarisation of the emitted light is lost is related to the tumbling rate of the molecule and hence its size. When the protein forms a large complex, the loss of polarisation is less. Thus, monitoring the change in polarisation can be used to detect the extent that interactions between the protein and other species occur. FP has been used to monitor the interactions between the RNase S protein and the RNase S peptide using a GFP reporter[97] and to help screen small molecule inhibitors of the p53-DM2 protein complex[98]. While this technique is relatively simple, requiring only a single fluorophore, it too, is limited to *in vitro* conditions.

Several of the methods described above, can provide detailed information and help to map the protein interactome through the analysis of protein-protein interactions in a low or medium throughput manner. More detailed information about the kinetics and

thermodynamics can also be achieved through the analysis of two or more binding partners. However all of the methods described so far, are limited to *in vitro* conditions, which means that any information obtained is done so in an artificial environment. Methods that can detect protein-protein interactions both *in vitro* and *in vivo* would provide the highest utility for understanding the role protein oligomerization plays in biology.

1.6.4. Split protein approach

Richards in 1957 demonstrated that when the enzyme bovine pancreatic ribonuclease A (RNAase A) is digested with the protease subtalisin, it is cleaved into two parts, RNAase S protein and RNAase S peptide, that individually have no activity but when mixed and refolded correctly, retain catalytic function[99]. In a conceptually similar manner, several proteins can be split into different parts that retain their function only when reassembled (Figure 1.5). These systems can be exploited to monitor protein-protein interactions if each pair of the split protein are fused to each of the protein partners of interest. This can be achieved both *in vitro* and importantly, *in vivo*.

1.6.4.1. Split Green fluorescent protein

Green fluorescent protein (GFP) and its derivatives are among the most highly used protein-based tools in biochemistry and cell biology [100, 101] and the development of GFP earned the Nobel Prize in 2008[102, 103, 104]. GFP is the β-barrelled protein responsible for the green fluorescence of the *Aequorea victoria* species of jellyfish. Additionally, it has been engineered for optimised stability, increased quantum yield[105, 106, 107, 108] and to produce many different colours such as yellow (YFP)[109], blue (BFP)[110] and cyan (CFP)[111].

While originally used as a localisation tag or transcriptional reporter, GFP has also been engineered to detect protein-protein interactions. This has been achieved through “split”-GFP as a non-enzymatic analogy to RNAase A described above (see section 1.6.4.) [112, 113]. Each half of GFP or a derivative which are non-fluorescent individually, are fused to each of the proteins

of interest that are thought to interact, with their interaction being confirmed by the recovery of fluorescence. Split GFP has also been used to improve the expression and purification of membrane proteins through comparison of active, soluble protein and inactive proteins observed in inclusion bodies[114]. Furthermore, split GFP can be used to localise protein interactions inside yeast cells[115]. Related to the research described in this thesis split GFP has been used to report on the assembled structure of the papillomavirus capsid[116].

1.6.4.2. Split Luciferase

Firefly Luciferase converts luciferin to oxyluciferin which is bioluminescent [117]. Similar to RNAase, when split, Luciferase loses its enzymatic activity, however, upon reassembly, this activity is restored. In split mode, Luciferase is a very common *in vivo* method to detect protein-protein interactions [118, 119]. To date, split luciferase has been used for applications to monitor the formation of DnaE in mammalian cells[120], and to study the interactions between insulin and transforming growth factor beta (TGF-β)[121].

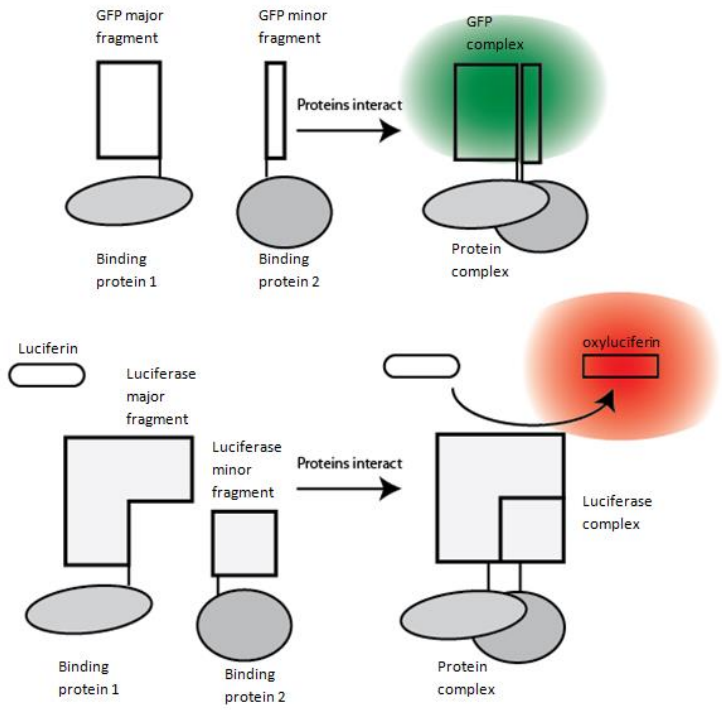


Figure 1.5 – Cartoon representations of "split protein" approaches to protein-protein interaction detection. (top) Split GFP (rectangles) reporter showing the restoration of fluorescence once the fluorescent protein is reassembled upon interaction between the proteins (ovals) to which each part of

GFP are fused. (bottom) Split luciferase reporter technology: When the proteins of interest interact (ovals), they bring together the two halves of the split luciferase (rectangles) which can then convert luciferin to bioluminescent oxyluciferin.

Conceptually these techniques use the formation of protein quaternary structure to observe protein quaternary structure. While able to expand the detection of protein-protein interactions into cells, these methods require large protein fusions which can alter the native folding, expression, and interaction of the target proteins.

1.6.5. Förster resonance energy transfer (FRET)

Theodor Förster discovered that energy can be transferred in a non-radiative manner between two proximal fluorophores. This property is referred to as Förster resonance energy transfer (FRET)[122]. When one of the fluorophores is excited, energy from the excited state can be transferred to the second fluorophore, which is then emitted at a wavelength unique to this fluorophore. This can only happen when the fluorophores are close in space. When they are distal, only the first fluorophore emits. Because FRET can identify the close proximity of two fluorophores, it can be used to detect, protein-protein interactions if the FRET donor and acceptor are placed on each of the binding partners[123, 124]. FRET has been used extensively for the study of protein interactions both inside and outside of cells. GFP and its derivatives are frequently used for FRET assays[125]. FRET has been used to study the activity of G-protein coupled receptors[126], the interaction of epidermal growth factor receptor with the Src homology 2 domain[127], as well as activation of the Fas receptor[128].

1.6.6. Two-hybrid screening

Another technique that has been widely used and that shares similarities to split protein assays, is two-hybrid screening, which was originally described in yeast cells[129] (hence “yeast two hybrid screening” or Y2H). This is one of the most widely used methods for the detection of protein-protein interactions inside of cells. This method works by using transcription factors which can exhibit different domains, where one domain binds to DNA and

the other actives transcription. By splitting these domains onto two proteins that are thought to interact, a detection system can be created. The first protein which is attached to the DNA binding domain locates to the section of DNA responsible for gene activation. If the second protein is able to bind to the first, it does so, thus bringing the activation domain into close proximity with its partner and switching on gene expression (Figure 1.6). This can lead to either a coloured response (expression of a fluorescent protein) or even cell survival if grown in nutrient lacking media.

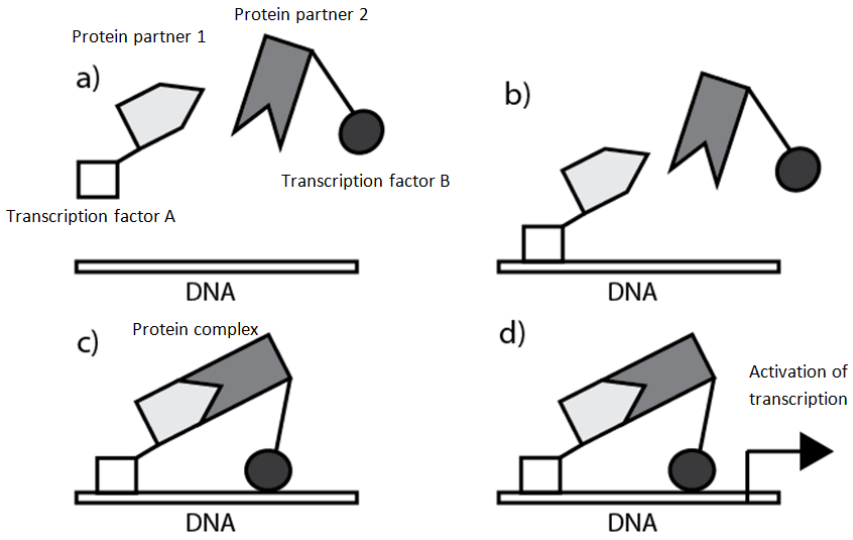


Figure 1.6 – Cartoon representation of two hybrid screening. a) Expression of two proteins each from a different plasmid, each fused to a different part of a transcription factor. b) Binding of the DNA transcription factor domain (square) to the DNA strand, c) Protein partners interact, bring the activation domain (circle) into close proximity to the DNA, d) which switches on the transcription of the gene of interest.

One advantage of two hybrid screening is that by using advanced cloning techniques, this method can be expanded to not just detect the interactions between two protein partners, but to map entire interactomes. This method has to date been used to map protein interactions for yeast[130], human cells[131], *E.coli*[132] and even Drosophila[133]. This ability to create libraries over large numbers of protein targets makes two hybrid screening a power tool for both fundamental science and for drug development. It could also be employed to screen libraries of protein designs to evaluate their binding potential to specific partners in a high

throughput manner. However, a high false positive rate is often observed providing some limitations or at least the need for robust secondary screens.

1.6.7. Fluorescein arsenical hairpin binder (FlAsH) and its conceptual origins in GFP.

GFP, as described, above is an invaluable tool in either its split form, when used for FRET, or for simple protein localisation experiments, but it is far from ideal. For example, the large size of GFP can lead to changes in the function or transport of the protein to which it is fused[134, 135]. To bypass these problems, Tsien and co-workers designed small chemical probes that can function as a GFP alternative for protein labelling. “Fluorescein arsenical hairpin binder” (FlAsH) and its alternative “resorufin arsenical hairpin binder” (ReAsH), were developed by positioning two arsenic atoms on the fluorophores fluorescein or resorufin. When the arsenic atoms are ligated to ethanedithiol (EDT), the molecules are weakly fluorescent[89, 136, 137]. When the EDT ligands are displaced by four sulfur atoms originating from cysteine residues positioned in an ideal geometry, the observable fluorescence intensity for these reagents greatly increases most likely due to a change in rotational properties about the carbon-arsenic bond (Figure 1.7)[138]. A short peptide including the four cysteines can be attached to any protein in analogy to GFP fusions, and, because the presences of four cysteines with this spacing are extremely rare, it provides orthogonality to the rest of the proteome. The ideal presentation of the four cysteines was initially thought to be a single face of a α-helix, but further optimisation resulted in the hairpin peptide CCPGCC[139, 140, 141, 142]. A family of arsenical hairpin binders has since been produced which not only vary in colour but in binding conformation as well,[89, 143] and this family of probes has been applied to protein aggregation[144], amyloid formation[145], tyrosine kinase activity[146] and protein purification[147].

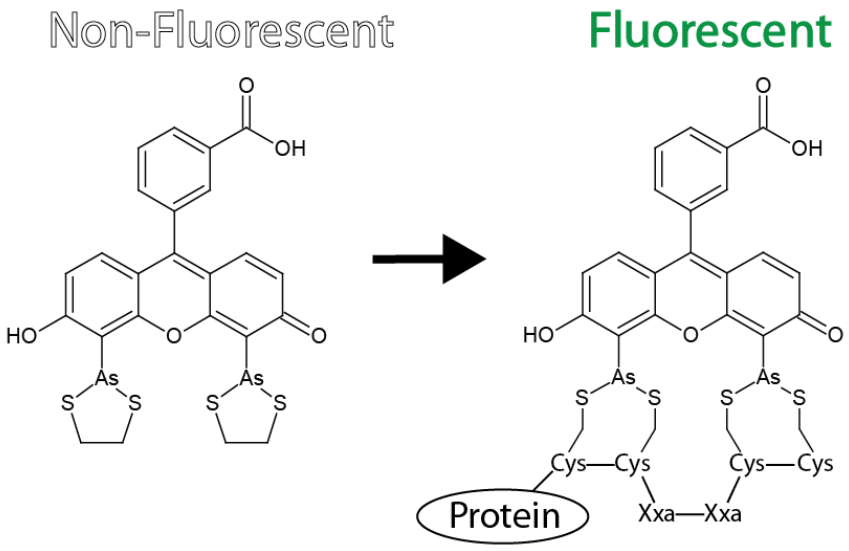


Figure 1.7 – The fluorescent activation of FlAsH-EDT$_2$. (left) Non fluorescent form of FlAsH bound to two EDT ligands, (right) Fluorescence is activated once these ligands are displaced by four sulfur atoms presented by four cysteine residues in an appropriate geometry.

FlAsH technology and the peptide sequence CCPGCC has been used to great effect but initially could be used to detect only single species. In response to this limitation, FlAsH was expanded for protein-protein detection applications. Tsien and co-workers eventually discovered peptide sequences that has even higher affinities for FlAsH/ReAsH than CCPGCC, and even found sequences that were more selective for one fluorophore over the other[142]. These sequences were used in a subsequent investigation to study the interactions between β-arrestin2 and its cell surface receptor PTH, by labelling each of the proteins with a different biarsenical fluorescent probe by using the peptide sequences with specificity for either FlAsH or ReAsH, and monitoring FRET once they interacted[148]. Cross-linked analogues of FlAsH-EDT$_2$ have been used by Schultz and co-workers to probe for protein-protein interactions inside cells [149]. By expressing two proteins in cells that each contain the ideal binding sequence, an interaction can be detected via addition of the cross-linked analogue. This provides an interesting method for the detection of protein-protein interactions.

An alternative approach was developed by Schepartz and co-workers, following on their realisation that the binding motif (CCPGCC) can be split into two 'bipartite' cysteine pairs

that, while distal in sequence, are proximal in space[150]. They demonstrated that bipartite cysteine pairs could be positioned across a protein-protein interface so long as the two sets of cysteines can collectively form a FlAsH-EDT_2 binding site[138, 150, 151]. The deconstruction of the ideal binding site into cysteine residues placed appropriately in space has allowed for the detection of interacting proteins using very small disruptions to the protein primary sequence and, as FlAsH-EDT_2 is cell permeable, it has been used to study the interactions of proteins *in vivo* in a structural context that is very near to native. Thus, the potential of such a tag, which can be easily appended to a protein and does not require complex post-translational modifications or large protein fusions, in concert with a small, cell-permeable fluorophore, cannot be understated. This technique will be the basis of our screen to assess the assembly of protein nanocages *in vitro* under different conditions, as described in Chapter 4 and 5, and will be scaled up to high throughput screens of protein libraries *in vivo*, as chronicled in Chapter 6.

1.6.7.2. Development of FlAsH-EDT_2 binding sites based on idealised geometry analysis

While termini do provide easy and useful locations for the formation of a binding site, it does limit the number of designs that can be formed, preventing this technology from being expanded to proteins that do not exhibit solvent accessible termini or protein ensembles without proximal termini. An investigation in 2009 by Goodman et al[152], evaluated the ability to expand FlAsH-EDT_2 technology to other types of secondary structures including unstructured strands, loops, beta sheets and a combination of the above, with an emphasis on moving the design process away from termini locations (Figure 1.8a)[138]. The proteins p53 and EmGFP were engineered to incorporate cysteines in various positions but on the principle of obeying geometry rules as stated by Shaikh et al[153], who analysed the bond angles of different arsenic dithiolates. As FlAsH-EDT_2 contains two sets of arsenic dithiolate bonding motifs, this investigation provides useful insights into the necessary geometric requirements for the relative positioning of the sulfur atoms, derived from the cysteine residues, to allow FlAsH binding (Figure 1.8b). This design information, as shown by Goodman et al[152], could be used to

guide FLAsH-EDT$_2$ binding site design development. Work performed on p53 and EmGFP showed that binding sites could be achieved on both β-sheets and loops. By taking into account the distance between the two arsenic atoms on FlAsH, an ideal geometric limitation, in the form of a rhombus of 4 – 5 Å on one side and 6 – 8 Å on the other, was established. An interesting note is that this strategy is similar to the original FlAsH-EDT$_2$ binding site, along the length of an α-helix as initially described by Tsien and co-workers[136], before the wide spread adoption of the peptide CCPGCC[140] and its analogues. However the expansion of this technique to incorporate not just multiple secondary structures but also multiple proteins across protein-protein interfaces would provide the highest utility of this technique for the protein nanocage field.

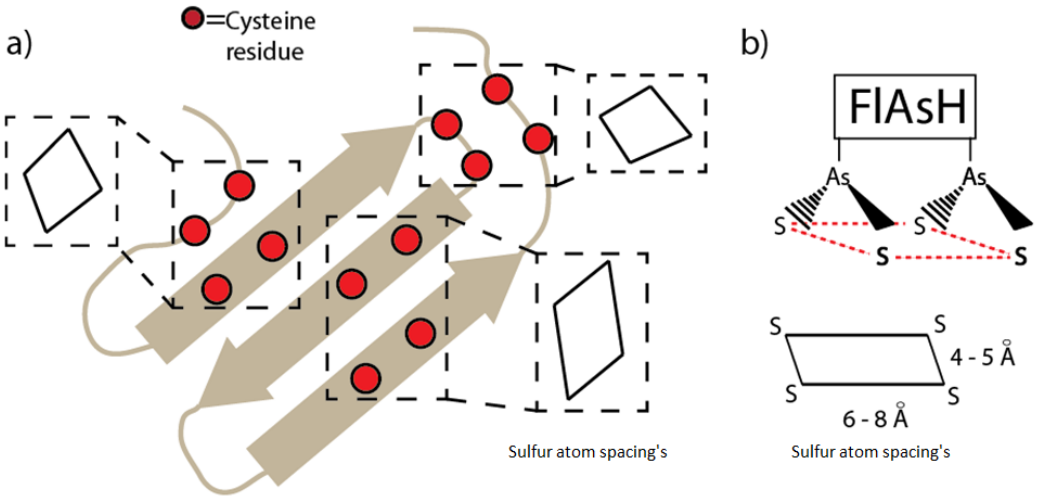


Figure 1.8 – a) Cartoon showing the advances by Schepartz and co-workers in establishing new FlAsH-EDT$_2$ binding sites (β-strands and loops) and the resulting rhombus shaped binding sites (red dots – cysteine residues). b) Proposed rhombus binding site geometric requirements based on arsenic dithiolate geometries.

The geometric design strategy described above was used to great effect by Scheck et al[154], when trying to unravel the conformational changes during epidermal growth factor receptor (EGFR) signal transduction through a membrane[155]. This study showed the power of geometrically driven, multi-protein, FlAsH/ReAsH binding site design to answer a fundamental

cell biological question. A similar strategy is adopted in part of the research described in Chapter 5 of this thesis.

1.6.7.3. FlAsH technology in living cells

To have the highest utility, any protein-protein interaction detection assay must be able to to work *in vitro* and *in vivo*. FlAsH-EDT_2 has been highly employed in mammalian cells[89, 156, 137, 140, 157] since its inception, as well as yeast cells[158] however its uses have been limited in bacterial cells[159]. Some studies have been undertaken to expand its utility into other cell types. Work performed by Enninga *et al*, followed the bacterial secretion of IpaB and IpaC as they came into contact with mammalian cells, via FlAsH-EDT_2 labelling[160, 161]. This study established methods to label proteins with cysteine rich tags inside bacterial cells and examine the effect of FlAsH-EDT_2. It was found that even high concentrations of FlAsH-EDT_2 (20 μM) had little effect on the vitality of the cell. This investigation showed that simple diffusion was enough to label proteins inside bacteria. Another study by Ignatova et al[144], labelled bacterial proteins and monitored the stability between their folded and unfolded states in bacterial cells. Their labelling technique required the short incubation of the cells with lysozyme in order to permeabilize the cells. Using this method to internalise FlAsH-EDT_2, membrane protein structure was investigated[162, 163]. The cellular stability of the MtrC was establishing using a combination of FlAsH-EDT_2 intercellular labelling and a cell-impermeable analogue, CrAsH)[164]. Although there are some published methods for the use of FlAsH-EDT_2 in bacterial cells, further work would be needed to optimise their use for an *in vivo* nanocage detection library assay. This was developed in the research described in Chapter 6.

1.7. Approaches for the engineering of novel protein-protein interactions

One strain of the research described in this thesis (Chapters 3 and 6) is to develop new protein nanocages with new properties and are thus conceptually synthetic in nature. As detailed above, many methods have been created to detect the interactions between proteins.

These methods have uses as wide as assessing protein binding to a partner, or in mapping an interactome. Tools like these to understand natural phenomena are invaluable to the scientific community and can aid knowledge of how proteins fulfil their function as well as in diagnosing diseases when their functions go awry. However, the ability to assess protein-protein interactions can also be used for non-natural protein applications.

The field of synthetic biology aims to create novel alternatives to what already exists in Nature, thus enveloping multiple fields of research under a broad umbrella. This breadth is best demonstrated through examples. One example of this is the engineering of enzymes to provide new catalysts for the chemical industry[165]. Another example is determining whether it is possible to create folded proteins from a smaller set of amino acids, as described by Hilvert and co-workers[166]. A third example described by Venter and co-workers[167] showed that an organism can be successfully encoded by a completely synthetic genome. A fourth example is Gellman and co-workers'[168] development of extended backbone amino acids to generate polyamides that robustly fold into novel tertiary structure.

A large area of synthetic biology research is to engineer proteins with new function. However, this research is slowed due to the very limited knowledge about how secondary structure leads to tertiary structure and then on to quaternary structure. High resolution crystal and NMR structures, while incredibly useful, provide only a static picture and little insight into mechanisms of folding or assembly. These are still difficult to predict, leaving the *de novo* design of protein structure as mostly unattainable.

1.7.1. Rational design and computational approaches to the engineering of protein-protein interactions

Research to analyse the propensity of amino acids to form certain secondary structures and the resulting development of secondary structure prediction software, were the starting points to create peptide sequences that could fold and interact. DeGrado and co-

workers showed that by using these tools, a 4-helix bundle, not dissimilar to that observed in ferritin monomers, could be formed from completely designed, synthetic peptides[169, 170]. This synthetic 4-helix bundle was later used by Dutton and co-workers who modified the bundle so that it was able to incorporate a heme group. This new design was able to act as a rudimentary version of myoglobin for oxygen storage, impressive because it was constructed of only a simple peptide folding sequence[171, 172]. This design demonstrated that not only can protein tertiary structure be predicted and designed, but that this simple design can be used as a scaffold for grafting similar functions to those seen in nature without the need for the more complex structure of Nature-derived proteins.

Woolfson and co-workers established design rules to expand the bundle from 4-helices to 6. Using semi-empirical methods, Woolfson and co-workers modified the peptide structure to extend the hydrophobic region that originally made up the core of the 4-helix bundle, to reposition the 'knob in hole' interactions to the edges of the bundle. This combination of changes resulted in a designed peptide sequence which could assemble into a novel 6-helix bundle with a central pore[173]. Woolfson and co-workers have expanded this technique to create arrays of peptides by controlling the hydrophobicity of different faces of the α-helix. This allowed for the construction of large nanostructures made up of hundreds of individual peptides[174], marking an introduction into large artificial nanostructure design.

While semi-empirical design has expanded the synthetic toolbox of protein folds, the number of designs that can be examined is limited. Using computational approaches, which would screen all possible conformations and are not subjected to the constraints of wet laboratory experiments, relevant designs can be virtually screened to fit set criteria.

Computational design has been used to engineer enzymatic function. Baker and co-workers used high powered computing technology to screen libraries of protein folds that could be manipulated to incorporate the necessary binding pocket needed to stabilise a

transition state during enzymatic catalysts. By virtually combining designed cavities with protein folds that had no native enzymatic function, enzymes that catalyse retro-aldol[175] and Kemp elimination reactions were generated[176]. By screening native protein folds with the computational approach, a protein was discovered that could bind to hemaglutinin with nanomolar affinity[177] highlighting the ability of this method to generate novel protein-protein interactions.

Recently, Baker's approach has been applied to the design of proteins that assemble into nanocages.[178] Building blocks were derived from proteins that natively provided the appropriate symmetries for the generation of key axes in the assembled nanocage. Two protein designs were found that could assemble with either tetrahedral or octahedral symmetries, similar to the mini- and maxi-ferritins described above. These reports emphatically demonstrated how far the design of quaternary structure has come, however, most of this research required the expression and characterisation of a large number of proteins to find those that actually conformed to the conformational design. This fact underlines an incongruity between computational designs and reality. Clearly, these need to be refined and a deeper understanding needs to be achieved to direct the refinement.

These computational approaches to screen "virtual libraries" of protein nanocages often require extensive computing power and can be met with a high error rates thus requiring extensive traditional 'wet lab' characterisation of the expressed and purified proteins *in vitro*. An alternative approach to computational screens is using "actual libraries" of protein mutants and a direct detection assay screen. This approach is conceptually similar to computational screens as they both aim to create a large library of different designs which are individually analysed for a specific attribute, such as nanocage formation. Because the screen of library members is basically a high throughput characterisation assay, it should limit the number of designs that require full characterisation, bypassing a weakness in the computational

technique. In addition, it could be imagined that a high throughput assay to directly assess nanocage formation could be coupled to the computational approach and help cull down the number of designs requiring full characterisation.

1.8. Cloning techniques for the generation of protein libraries

Chapter 6 of this thesis details the development of a high throughput assay to screen protein nanocage libraries. As part of that research, we were required to generate the libraries themselves. We therefore considered the several methods that have been developed to create large mutant protein libraries.

1.8.1. Error prone PCR and DNA shuffling: Generation of diversity across entire genes or plasmids

A number of techniques exist to create diversity in proteins by generating diversity in the DNA that encodes them. One method to generate diversity is through error prone PCR where the PCR buffer conditions are manipulated and a low fidelity enzyme Taq polymerase is used, This results in the mis-incorporation of bases into the PCR product[179]. Mutations in the polymerase[180], changing the concentration of $MgCl_2$ and $MnCl_2$[181] or unequal use of the four dNTP's affect the mutation rate and allow a high level of control[182, 183, 184]. This technique is somewhat imprecise in that it generates diversity in random locations between the PCR primers, but it has since been used to great effect for directed evolution where mutagenesis in implemented between screening steps.

Error prone PCR was used by Arnold and co-workers to enhance the stability and function of the enzyme sublitisin E, greatly increasing its efficiency and survivability in high concentrations of dimethylformamide[185, 186]. This method has been used for other enzymes[187] and was even employed to evolve enzymes to accept and turnover new substrates. Random libraries of Cytochrome P450 monooxygenase were created and screened to select for the ability to oxidise alkanes with a higher efficiency over its normal fatty acid substrates. By

performing multiple rounds of mutation with selection cycles in between, the enzyme mutants were directed towards this new function[188, 189]. Error prone PCR has also been used for directed evolution of a protein container to sequester HIV protease[74]. Together, these and similar studies demonstrated that, with clever screen design, protein function can be altered toward new and novel applications without loss in protein stability, loss of solubility and without the need for extensive screening out of false positives as seen in "virtual screening".

DNA shuffling can provide either a complementary or alternative method to error prone PCR[190]. One limitation of error prone PCR is that even successful mutations appear only a minimal number of times. Hence, combinations of beneficial mutations are rare. DNA shuffling, which involves the random digestion of a gene library, followed by the random reassembly of overlapping fragments through PCR. Repeated rounds of digestion and error prone PCR can shuffle up the mutations combinatorially. The resulting libraries can then be used in directed evolution screens [107].

While both of these techniques are very useful for the creation of mutant libraries, there is little control over what position mutations occur. They can occur between the amplification primers and hence the points of diversification are inaccurately placed. This fact restricts the structural precision of a diversification strategy raising the possibility that a legitimate "hit" library member will have "success" through an undesired mechanism. This could be especially problematic if the screening step indirectly assesses a desired property (ie a survivability screen to find enzymes with a beta barrel fold). In addition, many of the library members may have limited activity due to protein misfolding thus wasting diversity. These artefacts can be compounded if error prone PCR is used to amplify the whole plasmid; copy number, origin sites and protein expression levels could be altered thereby lending non-screening controlled bias to the library. Library generation techniques that provide precise

control of the placement of mutations give greater control of the diversity elements in the protein, and more useful diversity space is explored.

1.8.2. Transfer PCR (TPCR) and saturation mutagenesis: Diversity at specific positions

Error prone PCR is a frequently used method for the generation of mutant protein libraries; however it is limited by being unselective in its placement of mutations. A novel technique takes a similar strategy to site directed mutagenesis and has been used for the creation of large mutant libraries is transfer PCR (TPCR)[191, 192]. This method originally described by Peleg and co-workers, allows for the rapid incorporation of several site directed mutations on a single gene in a one pot reaction. It uses several forward primers containing the desired mutations and only one reverse primer. The PCR is split into two parts, where first the primers generate super primers of different lengths along the gene including a collection with the desired mutations. The second stage incorporates these super primers back into the plasmid (Figure 1.9). While Peleg and co-workers used TPCR to create a single mutant with multiple mutations, it is straightforward to understand how a library could be generated if the mutated codons were randomised. Therefore, precise positions can be randomised, multiple positions can be randomised, and randomised positions can be coupled together on a single primer. A consequence of this is that it is possible to limit the area for the random incorporation of amino acids to certain structural regions of the protein, such as at enzyme active sites, at specific surfaces or at a single protein-protein interface.

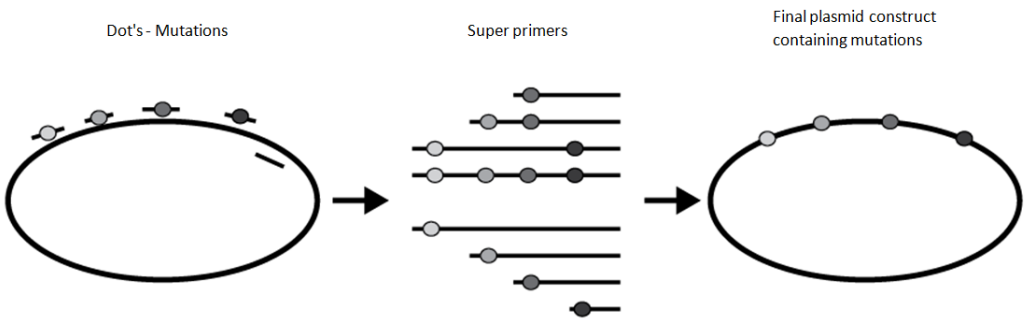


Figure 1.9 – Schematised explanation of the two step transfer PCR (TPCR) protocol that was used to generate protein nanocage libraries in Chapter 6. (Left) Circular plasmid containing the gene of interest with the locations of multiple forward primers containing different sites of random mutation (grey dots) and a single reverse primer. (Middle) The first step of the TPCR reaction creates multiple sized "super primers" that contain at least one, of the mutations. (Right) The second step of TPCR incorporates the "super primers" into the gene of interest in the plasmid.

Since its inception, TPCR has been used to add restriction sites for fast cloning methods[193], create mutant kinases[194], and explore the binding landscapes of protein interactions[195]. Shifman and co-workers using ORBIT[196] computational screening, a method similar to Baker's (see above) increased the binding affinity between two proteins that only originally weakly interacted[197]. The predicted hits for characterisation were generated by TPCR. However, Shifman and co-worker's paper, only envisioned the use of TPCR for the creation of large random mutant libraries and this was not the aim of the original study.

A technique that shares several similarities to TPCR, is saturation mutagenesis[198, 199]. This method selects only a few residues for site directed mutagenesis with degenerate codon primers allowing for complete saturation of all possibilities onto those particular sites. This method has been used to alter the accessibility of substrates for the enzyme transketolase[200, 201, 202] as well as enhancing its stereoselectivity[203].

TPCR provides advantages to error prone PCR by limiting the positions that are randomised. This increases the useable diversity in the library and reduces the occurrence of mutations at spurious positions that can affect have artefact generating consequences. However, it should be noted that the chance of finding a mutation that can provide a desired effect but through an unexpected mechanism would be limited. As a full understanding of

protein folding is lacking, it might be these 'unexpected' results that could lead to the largest breakthroughs. We will discuss our use of TPCR in Chapter 6 to create libraries of a protein with crippled nanocage forming ability. We will develop a high throuhgput screen of these libraries to find novel proteins with recovered oligomerization.

1.9. Goals of this investigation

In the research chronicled in this thesis, we aim to contribute to the fundamental understanding of the role protein-protein interactions play in the assembly of ferritin nanocage proteins, to the exploitation of these proteins in nanotechnology, and to the development of new tools to advance this field. We begin by exploring the application of ferritins to the generation of gold nanoparticles formation and to targeted delivery (Chapter 2). These and other applications will be enhanced by beginning to establish an understanding of symmetry axis formation and of the importance played by extra-bundle helicies to the assembly of ferritin nanocages through crystallography (Chapter 3). Furthermore we will develop a direct assay for ferritin nanocage formation, based on FlAsH technology that will allow the rapid screening of multiple conditions to optimise the nanoparticle generating reactions (Chapters 4 and 5). Finally, we will further develop this assay for screening, in live cells, libraries of ferritins with randomised protein-protein interfaces to enhance protein nanocage assembly (Chapter 6).

References

1. Whitesides, G. M.; Grzybowski, B., Self-assembly at all scales. *Science* **2002,** *295* (5564), 2418-2421.
2. Costerton, J. W.; Cheng, K. J.; Geesey, G. G.; Ladd, T. I.; Nickel, J. C.; Dasgupta, M.; Marrie, T. J., Bacterial biofilms in nature and disease. *Annual Review of Microbiology* **1987,** *41*, 435-464.
3. Zhang, S. G., Fabrication of novel biomaterials through molecular self-assembly. *Nature Biotechnology* **2003,** *21* (10), 1171-1178.
4. Janin, J.; Bahadur, R. P.; Chakrabarti, P., Protein-protein interaction and quaternary structure. *Quarterly Reviews of Biophysics* **2008,** *41* (2), 133-180.
5. Dill, K. A.; Ozkan, S. B.; Shell, M. S.; Weikl, T. R., The protein folding problem. In *Annual Review of Biophysics*, 2008; Vol. 37, pp 289-316.
6. Walter, S.; Buchner, J., Molecular chaperones - Cellular machines for protein folding. *Angewandte Chemie-International Edition* **2002,** *41* (7), 1098-1113.
7. Bromley, E. H. C.; Channon, K.; Moutevelis, E.; Woolfson, D. N., Peptide and protein building blocks for synthetic biology: From programming biomolecules to self-organized biomolecular systems. *Acs Chemical Biology* **2008,** *3* (1), 38-50.
8. Pawson, T.; Nash, P., Protein-protein interactions define specificity in signal transduction. *Genes & Development* **2000,** *14* (9), 1027-1047.
9. Prives, C.; Hall, P. A., The P53 pathway. *Journal of Pathology* **1999,** *187* (1), 112-126.
10. Dalrymple, B. P.; Kongsuwan, K.; Wijffels, G.; Dixon, N. E.; Jennings, P. A., A universal protein-protein interaction motif in the eubacterial DNA replication and repair systems. *Proceedings of the National Academy of Sciences of the United States of America* **2001,** *98* (20), 11627-11632.
11. Halliburton, W. D., On Muscle-Plasma. *The Journal of physiology* **1887,** *8* (3-4), 133-202.
12. Pollard, T. D.; Borisy, G. G., Cellular motility driven by assembly and disassembly of actin filaments. *Cell* **2003,** *112* (4), 453-465.
13. Pellegrini, F.; Budman, D. R., Review: Tubulin function, action of antitubulin drugs, and new drug development. *Cancer Investigation* **2005,** *23* (3), 264-273.
14. Selkoe, D. J., Alzheimer's disease: Genes, proteins, and therapy. *Physiological Reviews* **2001,** *81* (2), 741-766.
15. Vivekanandan, S.; Brender, J. R.; Lee, S. Y.; Ramamoorthy, A., A partially folded structure of amyloid-beta(1-40) in an aqueous environment. *Biochemical and Biophysical Research Communications* **2011,** *411* (2), 312-316.
16. Prusiner, S. B., Prions. *Proceedings of the National Academy of Sciences of the United States of America* **1998,** *95* (23), 13363-13383.
17. Bogan, A. A.; Thorn, K. S., Anatomy of hot spots in protein interfaces. *Journal of Molecular Biology* **1998,** *280* (1), 1-9.
18. Ellis, J., Porcine Circovirus A Historical Perspective. *Veterinary Pathology* **2014,** *51* (2), 315-327.
19. Legendre, M.; Bartoli, J.; Shmakova, L.; Jeudy, S.; Labadie, K.; Adrait, A.; Lescot, M.; Poirot, O.; Bertaux, L.; Bruley, C.; Coute, Y.; Rivkina, E.; Abergel, C.; Claverie, J.-M., Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. *Proceedings of the National Academy of Sciences of the United States of America* **2014,** *111* (11), 4274-4279.
20. Zandi, R.; Reguera, D.; Bruinsma, R. F.; Gelbart, W. M.; Rudnick, J., Origin of icosahedral symmetry in viruses. *Proceedings of the National Academy of Sciences of the United States of America* **2004,** *101* (44), 15556-15560.

21. Burnett, R. M., The structure of the adenovirus capsid.2. the packing symmetry of hexon and its implications for viral architecture. *Journal of Molecular Biology* **1985,** *185* (1), 125-143.
22. Yamada, S.; Matsuzawa, T.; Yamada, K.; Yoshioka, S.; Ono, S.; Hishinuma, T., Modified inversion recovery method for NMR imaging *Science Reports of the Research Institutes Tohoku University Series C Medicine* **1986,** *33* (1-4), 9-15.
23. Ohtsuka, K.; Hata, M., Molecular chaperone function of mammalian Hsp70 and Hsp40 - a review. *International Journal of Hyperthermia* **2000,** *16* (3), 231-245.
24. Zhang, X.; Konarev, P. V.; Petoukhov, M. V.; Svergun, D. I.; Xing, L.; Cheng, R. H.; Haase, I.; Fischer, M.; Bacher, A.; Ladenstein, R.; Meining, W., Multiple assembly states of lumazine synthase: A model relating catalytic function and molecular assembly. *Journal of Molecular Biology* **2006,** *362* (4), 753-770.
25. Kedersha, N. L.; Heuser, J. E.; Chugani, D. C.; Rome, L. H., Vaults.3. Vault ribonucleoprotein-particles open into flower-like structures with octagonal symmetry. *Journal of Cell Biology* **1991,** *112* (2), 225-235.
26. Tanaka, H.; Kato, K.; Yamashita, E.; Sumizawa, T.; Zhou, Y.; Yao, M.; Iwasaki, K.; Yoshimura, M.; Tsukihara, T., The Structure of Rat Liver Vault at 3.5 Angstrom Resolution. *Science* **2009,** *323* (5912), 384-388.
27. Dautant, A.; Meyer, J. B.; Yariv, J.; Precigoux, G.; Sweet, R. M.; Kalb, A. J.; Frolow, F., Structure of a monoclinic crystal form of cytochrome b1 (bacterioferritin) from E-coli. *Acta Crystallographica Section D-Biological Crystallography* **1998,** *54*, 16-24.
28. Zhang, X. F.; Meining, W.; Fischer, M.; Bacher, A.; Ladenstein, R., X-ray structure analysis and crystallographic refinement of lumazine synthase from the hyperthermophile Aquifex aeolicus at 1.6 angstrom resolution: Determinants of thermostability revealed from structural comparisons. *Journal of Molecular Biology* **2001,** *306* (5), 1099-1114.
29. Kaufmann, B.; Simpson, A. A.; Rossmann, M. G., The structure of human parvovirus B19. *Proceedings of the National Academy of Sciences of the United States of America* **2004,** *101* (32), 11628-11633.
30. Vu, N. T.; Moriwaki, Y.; Caaveiro, J. M. M.; Terada, T.; Tsutsumi, H.; Hamachi, I.; Shimizu, K.; Tsumoto, K., Selective binding of antimicrobial porphyrins to the heme-receptor IsdH-NEAT3 of Staphylococcus aureus. *Protein Science* **2013,** *22* (7), 942-953.
31. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **2004,** *25* (13), 1605-1612.
32. Sapsford, K. E.; Algar, W. R.; Berti, L.; Gemmill, K. B.; Casey, B. J.; Oh, E.; Stewart, M. H.; Medintz, I. L., Functionalizing Nanoparticles with Biological Molecules: Developing Chemistries that Facilitate Nanotechnology. *Chemical Reviews* **2013,** *113* (3), 1904-2074.
33. Kitagawa, T.; Kosuge, H.; Uchida, M.; Dua, M. M.; Iida, Y.; Dalman, R. L.; Douglas, T.; McConnell, M. V., RGD-Conjugated Human Ferritin Nanoparticles for Imaging Vascular Inflammation and Angiogenesis in Experimental Carotid and Aortic Disease. *Molecular Imaging and Biology* **2012,** *14* (3), 315-324.
34. Li, K.; Zhang, Z.-P.; Luo, M.; Yu, X.; Han, Y.; Wei, H.-P.; Cui, Z.-Q.; Zhang, X.-E., Multifunctional ferritin cage nanostructures for fluorescence and MR imaging of tumor cells. *Nanoscale* **2012,** *4* (1), 188-193.
35. Douglas, T.; Young, M., Viruses: Making friends with old foes. *Science* **2006,** *312* (5775), 873-875.
36. Theil, E. C., Ferritin - structure, gene-regulation, and cellular function in animals, plants, and microorganisms. *Annual Review of Biochemistry* **1987,** *56*, 289-315.

37. Hoare, R. J.; Harrison, P. M.; Hoy, T. G., Structure of horse-spleen apoferritin at 6-A resolution. *Nature* **1975,** *255* (5510), 653-654.
38. Crichton, R. R.; Declercq, J.-P., X-ray structures of ferritins and related proteins. *Biochimica Et Biophysica Acta-General Subjects* **2010,** *1800* (8), 706-718.
39. Andrews, S. C.; Arosio, P.; Bottke, W.; Briat, J. F.; Vondarl, M.; Harrison, P. M.; Laulhere, J. P.; Levi, S.; Lobreaux, S.; Yewdall, S. J., Structure, function, and evolution of ferritins. *Journal of Inorganic Biochemistry* **1992,** *47* (3-4), 161-174.
40. Fan, R.; Boyle, A. L.; Cheong, V. V.; Ng, S. L.; Orner, B. P., A Helix Swapping Study of Two Protein Cages. *Biochemistry* **2009,** *48* (24), 5623-5630.
41. Ardejani, M. S.; Chok, X. L.; Foo, C. J.; Orner, B. P., Complete shift of ferritin oligomerization toward nanocage assembly via engineered protein-protein interactions. *Chemical Communications* **2013,** *49* (34), 3528-3530.
42. Johnson, E.; Cascio, D.; Sawaya, M. R.; Gingery, M.; Schroder, I., Crystal structures of a tetrahedral open pore ferritin from the hyperthermophilic Archaeon Archaeoglobus fulgidus. *Structure* **2005,** *13* (4), 637-648.
43. Aisen, P.; Listowsky, I., Iron transport and storage proteins. *Annual Review of Biochemistry* **1980,** *49*, 357-393.
44. Yariv, J.; Kalb, A. J.; Sperling, R.; Bauminger, E. R.; Cohen, S. G.; Ofer, S., The composition and the structure of bacterioferritin of Escherichia-coli. *Biochemical Journal* **1981,** *197* (1), 171-175.
45. Hintze, K. J.; Theil, E. C., Cellular regulation and molecular interactions of the ferritins. *Cellular and Molecular Life Sciences* **2006,** *63* (5), 591-600.
46. Zhang, Y.; Raudah, S.; Teo, H.; Teo, G. W. S.; Fan, R.; Sun, X.; Orner, B. P., Alanine-shaving Mutagenesis to Determine Key Interfacial Residues Governing the Assembly of a Nano-cage Maxi-ferritin. *Journal of Biological Chemistry* **2010,** *285* (16), 12078-12086.
47. Ardejani, M. S.; Li, N. X.; Orner, B. P., Stabilization of a Protein Nanocage through the Plugging of a Protein-Protein Interfacial Water Pocket. *Biochemistry* **2011,** *50* (19), 4029-4037.
48. Ilari, A.; Stefanini, S.; Chiancone, E.; Tsernoglou, D., The dodecameric ferritin from Listeria innocua contains a novel intersubunit iron-binding site. *Nature Structural Biology* **2000,** *7* (1), 38-43.
49. Chiancone, E.; Ceci, P., The multifaceted capacity of Dps proteins to combat bacterial stress conditions: Detoxification of iron and hydrogen peroxide and DNA binding. *Biochimica Et Biophysica Acta-General Subjects* **2010,** *1800* (8), 798-805.
50. Zhang, Y.; Fu, J.; Chee, S. Y.; Ang, E. X. W.; Orner, B. P., Rational disruption of the oligomerization of the mini-ferritin E. coli DPS through protein-protein interface mutation. *Protein Science* **2011,** *20* (11), 1907-1917.
51. Grant, R. A.; Filman, D. J.; Finkel, S. E.; Kolter, R.; Hogle, J. M., The crystal structure of Dps, a ferritin homolog that binds and protects DNA. *Nature Structural Biology* **1998,** *5* (4), 294-303.
52. Theil, E. C., Coordinating responses to iron and oxygen stress with DNA and mRNA promoters: The ferritin story. *Biometals* **2007,** *20* (3-4), 513-521.
53. Chiancone, E.; Ceci, P.; Ilari, A.; Ribacchi, F.; Stefanini, S., Iron and proteins for iron storage and detoxification. *Biometals* **2004,** *17* (3), 197-202.
54. Bevers, L. E.; Theil, E. C., Maxi- and mini-ferritins: minerals and protein nanocages. *Progress in molecular and subcellular biology* **2011,** *52*, 29-47.
55. Li, L.; Fang, C. J.; Ryan, J. C.; Niemi, E. C.; Lebron, J. A.; Bjorkman, P. J.; Arase, H.; Torti, F. M.; Torti, S. V.; Nakamura, M. C.; Seaman, W. E., Binding and uptake of H-ferritin are mediated by human transferrin receptor-1. *Proceedings of the National Academy of Sciences of the United States of America* **2010,** *107* (8), 3505-3510.

56. Lim, S.; Ieee, Protein Cages as Theranostic Agent Carriers. *6th Ieee International Conference on Nano/Molecular Medicine and Engineering (Ieee-Nanomed 2012)* **2012**.
57. Harrison, P. M.; Arosio, P., Ferritins: Molecular properties, iron storage function and cellular regulation. *Biochimica Et Biophysica Acta-Bioenergetics* **1996,** *1275* (3), 161-203.
58. Baraibar, M. A.; Barbeito, A. G.; Muhoberac, B. B.; Vidal, R., Iron-mediated Aggregation and a Localized Structural Change Characterize Ferritin from a Mutant Light Chain Polypeptide That Causes Neurodegeneration. *Journal of Biological Chemistry* **2008,** *283* (46), 31679-31689.
59. Curtis, A. R. J.; Fey, C.; Morris, C. M.; Bindoff, L. A.; Ince, P. G.; Chinnery, P. F.; Coulthard, A.; Jackson, M. J.; Jackson, A. P.; McHale, D. P.; Hay, D.; Barker, W. A.; Markham, A. F.; Bates, D.; Curtis, A.; Burn, J., Mutation in the gene encoding ferritin light polypeptide causes dominant adult-onset basal ganglia disease. *Nature Genetics* **2001,** *28* (4), 350-354.
60. Butts, C. A.; Swift, J.; Kang, S. G.; Di Costanzo, L.; Christiansen, D. W.; Saven, J. G.; Dmochowski, I. J., Directing noble metal ion chemistry within a designed ferritin protein. *Biochemistry* **2008,** *47* (48), 12729-12739.
61. Zhang, L.; Swift, J.; Butts, C. A.; Yerubandi, V.; Dmochowski, I. J., Structure and activity of apoferritin-stabilized gold nanoparticles. *Journal of Inorganic Biochemistry* **2007,** *101* (11-12), 1719-1729.
62. Fan, R.; Chew, S. W.; Cheong, V. V.; Orner, B. P., Fabrication of Gold Nanoparticles Inside Unmodified Horse Spleen Apoferritin. *Small* **2010,** *6* (14), 1483-1487.
63. Keyes, J. D.; Hilton, R. J.; Farrer, J.; Watt, R. K., Ferritin as a photocatalyst and scaffold for gold nanoparticle synthesis. *Journal of Nanoparticle Research* **2011,** *13* (6), 2563-2575.
64. Butts, C. A.; Swift, J.; Kang, S.-g.; Di Costanzo, L.; Christianson, D. W.; Saven, J. G.; Dmochowski, I. J., Directing Noble Metal Ion Chemistry within a Designed Ferritin Protein. *Biochemistry* **2008,** *47* (48), 12729-12739.
65. Cheung-Lau, J. C.; Liu, D.; Pulsipher, K. W.; Liu, W.; Dmochowski, I. J., Engineering a well-ordered, functional protein-gold nanoparticle assembly. *Journal of Inorganic Biochemistry* **2014,** *130*, 59-68.
66. Kramer, R. M.; Li, C.; Carter, D. C.; Stone, M. O.; Naik, R. R., Engineered protein cages for nanomaterial synthesis. *Journal of the American Chemical Society* **2004,** *126* (41), 13282-13286.
67. Galvez, N.; Sanchez, P.; Dominguez-Vera, J. M., Preparation of Cu and CuFe Prussian Blue derivative nanoparticles using the apoferritin cavity as nanoreactor. *Dalton Transactions* **2005,** (15), 2492-2494.
68. Yamashita, I.; Hayashi, J.; Hara, M., Bio-template synthesis of uniform CdSe nanoparticles using cage-shaped protein, apoferritin. *Chemistry Letters* **2004,** *33* (9), 1158-1159.
69. Ueno, T.; Suzuki, M.; Goto, T.; Matsumoto, T.; Nagayama, K.; Watanabe, Y., Size-selective olefin hydrogenation by a Pd nanocluster provided in an apo-ferritin cage. *Angewandte Chemie-International Edition* **2004,** *43* (19), 2527-2530.
70. Zhang, L.; Laug, L.; Muenchgesang, W.; Pippel, E.; Goesele, U.; Brandsch, M.; Knez, M., Reducing Stress on Cells with Apoferritin-Encapsulated Platinum Nanoparticles. *Nano Letters* **2010,** *10* (1), 219-223.
71. Daniel, M. C.; Astruc, D., Gold Nanoparticles: Assembly, Supramolecular Chemistry, Quantum-Size-Related Properties, and Applications Toward Biology, Catalysis, and Nanotechnology. *Chemical Reviews* **2004,** *104* (1), 293-346.

72. Shenton, W.; Mann, S.; Colfen, H.; Bacher, A.; Fischer, M., Synthesis of nanophase iron oxide in lumazine synthase capsids. *Angewandte Chemie-International Edition* **2001,** *40* (2), 442-445.
73. Douglas, T.; Young, M., Host-guest encapsulation of materials by assembled virus protein cages. *Nature* **1998,** *393* (6681), 152-155.
74. Woersdoerfer, B.; Woycechowsky, K. J.; Hilvert, D., Directed Evolution of a Protein Container. *Science* **2011,** *331* (6017), 589-592.
75. Sana, B.; Johnson, E.; Sheah, K.; Poh, C. L.; Lim, S., Iron-based ferritin nanocore as a contrast agent. *Biointerphases* **2010,** *5* (3), FA48-FA52.
76. Schoonen, L.; van Hest, J. C. M., Functionalization of protein-based nanocages for drug delivery applications. *Nanoscale* **2014,** *6* (13), 7124-7141.
77. Ruder, W. C.; Lu, T.; Collins, J. J., Synthetic Biology Moving into the Clinic. *Science* **2011,** *333* (6047), 1248-1252.
78. Pierce, M. M.; Raman, C. S.; Nall, B. T., Isothermal titration calorimetry of protein-protein interactions. *Methods-a Companion to Methods in Enzymology* **1999,** *19* (2), 213-221.
79. Velazquez-Campoy, A.; Leavitt, S. A.; Freire, E., Characterization of protein-protein interactions by isothermal titration calorimetry. *Methods in molecular biology (Clifton, N.J.)* **2004,** *261*, 35-54.
80. Chiu, M. H.; Prenner, E. J., Differential scanning calorimetry: An invaluable tool for a detailed thermodynamic characterization of macromolecules and their interactions. *Journal of pharmacy & bioallied sciences* **2011,** *3* (1), 39-59.
81. Brandts, J. F.; Lin, L. N., Study of strong to ultratight protein interactions using differential scanning calorimetry. *Biochemistry* **1990,** *29* (29), 6927-6940.
82. Figeys, D.; McBroom, L. D.; Moran, M. F., Mass spectrometry for the study of protein-protein interactions. *Methods* **2001,** *24* (3), 230-239.
83. Gupta, P. K.; Rustgi, S., Mass spectrometry: An essential tool for genome and proteome analysis. *Indian Journal of Biotechnology* **2003,** *2* (1), 48-64.
84. Song, J.; Wang, Z.; Ewing, R. M., Integrated analysis of the Wnt responsive proteome in human cells reveals diverse and cell-type specific networks. *Molecular Biosystems* **2014,** *10* (1), 45-53.
85. Aebersold, R.; Mann, M., Mass spectrometry-based proteomics. *Nature* **2003,** *422* (6928), 198-207.
86. Homola, J.; Yee, S. S.; Gauglitz, G., Surface plasmon resonance sensors: review. *Sensors and Actuators B-Chemical* **1999,** *54* (1-2), 3-15.
87. Cui, X. Q.; Yang, F.; Sha, Y. F.; Yang, X. R., Real-time immunoassay of ferritin using surface plasmon resonance biosensor. *Talanta* **2003,** *60* (1), 53-61.
88. Santini, S.; Di Agostino, S.; Coppari, E.; Bizzarri, A. R.; Blandino, G.; Cannistraro, S., Interaction of mutant p53 with p73: A Surface Plasmon Resonance and Atomic Force Spectroscopy study. *Biochimica Et Biophysica Acta-General Subjects* **2014,** *1840* (6), 1958-1964.
89. Giepmans, B. N. G.; Adams, S. R.; Ellisman, M. H.; Tsien, R. Y., Review - The fluorescent toolbox for assessing protein location and function. *Science* **2006,** *312* (5771), 217-224.
90. Towbin, H.; Staehelin, T.; Gordon, J., Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets - procedure and some applications. *Proceedings of the National Academy of Sciences of the United States of America* **1979,** *76* (9), 4350-4354.
91. Salim, K.; Fenton, T.; Bacha, J.; Urien-Rodriguez, H.; Bonnert, T.; Skynner, H. A.; Watts, E., Oligomerization of G-protein-coupled receptors shown by selective co-immunoprecipitation. *Journal of Biological Chemistry* **2002,** *277* (18), 15482-15485.

92. Geva, G.; Sharan, R., Identification of protein complexes from co-immunoprecipitation data. *Bioinformatics* **2011,** *27* (1), 111-117.
93. Engvall, E.; Perlmann, P., Enzyme-linked immunosorbent assay (ELISA) quantitative assay of immunoglobulin-G. *Immunochemistry* **1971,** *8* (9), 871-&.
94. Lequin, R. M., Enzyme Immunoassay (EIA)/Enzyme-Linked Immunosorbent Assay (ELISA). *Clinical Chemistry* **2005,** *51* (12), 2415-2418.
95. Paltiel, A. D.; Weinstein, M. C.; Kimmel, A. D.; Seage, G. R.; Losina, E.; Zhang, H.; Freedberg, K. A.; Walensky, R. P., Expanded screening for HIV in the United States - An analysis of cost-effectiveness. *New England Journal of Medicine* **2005,** *352* (6), 586-595.
96. Jameson, D. M.; Seifried, S. E., Quantification of protein-protein interactions using fluorescence polarization. *Methods-a Companion to Methods in Enzymology* **1999,** *19* (2), 222-233.
97. Park, S.-H.; Raines, R. T., Fluorescence polarization assay to quantify protein-protein interactions. *Methods in molecular biology (Clifton, N.J.)* **2004,** *261*, 161-6.
98. Knight, S. M. G.; Umezawa, N.; Lee, H. S.; Gellman, S. H.; Kay, B. K., A fluorescence polarization assay for the identification of inhibitors of the p53-DM2 protein-protein interaction. *Analytical Biochemistry* **2002,** *300* (2), 230-236.
99. Richards, F. M., On the enzymic activity of subtilisin-modified ribonclease. *Proceedings of the National Academy of Sciences of the United States of America* **1958,** *44* (2), 162-166.
100. Shimomura, O.; Johnson, F. H.; Saiga, Y., Extraction, purification and properties of aequorin, a bioluminescent protein from luminous hydromedusan, aequorea. *Journal of Cellular and Comparative Physiology* **1962,** *59* (3), 223-&.
101. Prasher, D. C.; Eckenrode, V. K.; Ward, W. W.; Prendergast, F. G.; Cormier, M. J., Primary structure of the Aequorea-Victoria green-fluorescent protein. *Gene* **1992,** *111* (2), 229-233.
102. Chalfie, M., GFP: Lighting Up Life (Nobel Lecture). *Angewandte Chemie International Edition* **2009,** *48* (31), 5603-5611.
103. Shimomura, O., Discovery of Green Fluorescent Protein (GFP) (Nobel Lecture). *Angewandte Chemie International Edition* **2009,** *48* (31), 5590-5602.
104. Tsien, R. Y., Constructing and Exploiting the Fluorescent Protein Paintbox (Nobel Lecture). *Angewandte Chemie International Edition* **2009,** *48* (31), 5612-5626.
105. Heim, R.; Prasher, D. C.; Tsien, R. Y., Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proceedings of the National Academy of Sciences of the United States of America* **1994,** *91* (26), 12501-12504.
106. Cubitt, A. B.; Heim, R.; Adams, S. R.; Boyd, A. E.; Gross, L. A.; Tsien, R. Y., Understanding, improving and using green fluorescent proteins. *Trends in Biochemical Sciences* **1995,** *20* (11), 448-455.
107. Crameri, A.; Whitehorn, E. A.; Tate, E.; Stemmer, W. P. C., Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nature Biotechnology* **1996,** *14* (3), 315-319.
108. Cormack, B. P.; Valdivia, R. H.; Falkow, S., FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **1996,** *173* (1), 33-38.
109. Nagai, T.; Ibata, K.; Park, E. S.; Kubota, M.; Mikoshiba, K.; Miyawaki, A., A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nature Biotechnology* **2002,** *20* (1), 87-90.
110. Shaner, N. C.; Patterson, G. H.; Davidson, M. W., Advances in fluorescent protein technology. *Journal of Cell Science* **2007,** *120* (24), 4247-4260.
111. Rizzo, M. A.; Springer, G. H.; Granada, B.; Piston, D. W., An improved cyan fluorescent protein variant useful for FRET. *Nature Biotechnology* **2004,** *22* (4), 445-449.

112. Ghosh, I.; Hamilton, A. D.; Regan, L., Antiparallel leucine zipper-directed protein reassembly: Application to the green fluorescent protein. *Journal of the American Chemical Society* **2000,** *122* (23), 5658-5659.
113. Magliery, T. J.; Wilson, C. G. M.; Pan, W. L.; Mishler, D.; Ghosh, I.; Hamilton, A. D.; Regan, L., Detecting protein-protein interactions with a green fluorescent protein fragment reassembly trap: Scope and mechanism. *Journal of the American Chemical Society* **2005,** *127* (1), 146-157.
114. Rodriguez-Banqueri, A.; Kowalczyk, L.; Palacin, M.; Luis Vazquez-Ibar, J., Assessment of membrane protein expression and stability using a split green fluorescent protein reporter. *Analytical Biochemistry* **2012,** *423* (1), 7-14.
115. Barnard, E.; Timson, D. J., Split-EGFP screens for the detection and localisation of protein-protein interactions in living yeast cells. *Methods in molecular biology (Clifton, N.J.)* **2010,** *638*, 303-17.
116. Buck, C. B.; Cheng, N.; Thompson, C. D.; Lowy, D. R.; Steven, A. C.; Schiller, J. T.; Trus, B. L., Arrangement of L2 within the papillomavirus capsid. *Journal of Virology* **2008,** *82* (11), 5190-5197.
117. Dewet, J. R.; Wood, K. V.; Deluca, M.; Helinski, D. R.; Subramani, S., Firefly luciferase gene - structure and expression in mammalian-cells. *Molecular and Cellular Biology* **1987,** *7* (2), 725-737.
118. Jester, B. W.; Cox, K. J.; Gaj, A.; Shomin, C. D.; Porter, J. R.; Ghosh, I., A Coiled-Coil Enabled Split-Luciferase Three-Hybrid System: Applied Toward Profiling Inhibitors of Protein Kinases. *Journal of the American Chemical Society* **2010,** *132* (33), 11727-11735.
119. Stains, C. I.; Furman, J. L.; Porter, J. R.; Rajagopal, S.; Li, Y.; Wyatt, R. T.; Ghosh, I., A General Approach for Receptor and Antibody-Targeted Detection of Native Proteins Utilizing Split-Luciferase Reassembly. *Acs Chemical Biology* **2010,** *5* (10), 943-952.
120. Ozawa, T.; Kaihara, A.; Sato, M.; Tachihara, K.; Umezawa, Y., Split luciferase as an optical probe for detecting protein-protein interactions in mammalian cells based on protein splicing. *Analytical Chemistry* **2001,** *73* (11), 2516-2521.
121. Remy, I.; Michnick, S. W., A highly sensitive protein-protein interaction assay based on Gaussia luciferase. *Nature Methods* **2006,** *3* (12), 977-979.
122. Pietraszewska-Bogiel, A.; Gadella, T. W. J., FRET microscopy: from principle to routine technology in cell biology. *Journal of Microscopy* **2011,** *241* (2), 111-118.
123. Kramer, H. E. A.; Fischer, P., The Scientific Work of Theodor Forster: A Brief Sketch of his Life and Personality. *Chemphyschem* **2011,** *12* (3), 555-558.
124. Lohse, M. J.; Nuber, S.; Hoffmann, C., Fluorescence/Bioluminescence Resonance Energy Transfer Techniques to Study G-Protein-Coupled Receptor Activation and Signaling. *Pharmacological Reviews* **2012,** *64* (2), 299-336.
125. Sekar, R. B.; Periasamy, A., Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations. *Journal of Cell Biology* **2003,** *160* (5), 629-633.
126. Hoffmann, C.; Gaietta, G.; Bunemann, M.; Adams, S. R.; Oberdorff-Maass, S.; Behr, B.; Vilardaga, J. P.; Tsien, R. Y.; Eisman, M. H.; Lohse, M. J., A FlAsH-based FRET approach to determine G protein - coupled receptor activation in living cells. *Nature Methods* **2005,** *2* (3), 171-176.
127. Sorkin, A.; McClure, M.; Huang, F. T.; Carter, R., Interaction of EGF receptor and Grb2 in living cells visualized by fluorescence resonance energy transfer (FRET) microscopy. *Current Biology* **2000,** *10* (21), 1395-1398.
128. Siegel, R. M.; Frederiksen, J. K.; Zacharias, D. A.; Chan, F. K. M.; Johnson, M.; Lynch, D.; Tsien, R. Y.; Lenardo, M. J., Fas preassociation required for apoptosis signaling and dominant inhibition by pathogenic mutations. *Science* **2000,** *288* (5475), 2354-2357.

129. Fields, S.; Song, O.-k., A novel genetic system to detect protein–protein interactions. *Nature* **1989,** *340* (6230), 245-246.
130. Barabasi, A. L.; Oltvai, Z. N., Network biology: Understanding the cell's functional organization. *Nature Reviews Genetics* **2004,** *5* (2), 101-U15.
131. Rual, J. F.; Venkatesan, K.; Hao, T.; Hirozane-Kishikawa, T.; Dricot, A.; Li, N.; Berriz, G. F.; Gibbons, F. D.; Dreze, M.; Ayivi-Guedehoussou, N.; Klitgord, N.; Simon, C.; Boxem, M.; Milstein, S.; Rosenberg, J.; Goldberg, D. S.; Zhang, L. V.; Wong, S. L.; Franklin, G.; Li, S. M.; Albala, J. S.; Lim, J. H.; Fraughton, C.; Llamosas, E.; Cevik, S.; Bex, C.; Lamesch, P.; Sikorski, R. S.; Vandenhaute, J.; Zoghbi, H. Y.; Smolyar, A.; Bosak, S.; Sequerra, R.; Doucette-Stamm, L.; Cusick, M. E.; Hill, D. E.; Roth, F. P.; Vidal, M., Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **2005,** *437* (7062), 1173-1178.
132. Hu, J. C.; Kornacker, M. G.; Hochschild, A., Escherichia coli One- and Two-Hybrid Systems for the Analysis and Identification of Protein–Protein Interactions. *Methods* **2000,** *20* (1), 80-94.
133. Giot, L.; Bader, J. S.; Brouwer, C.; Chaudhuri, A.; Kuang, B.; Li, Y.; Hao, Y. L.; Ooi, C. E.; Godwin, B.; Vitols, E.; Vijayadamodar, G.; Pochart, P.; Machineni, H.; Welsh, M.; Kong, Y.; Zerhusen, B.; Malcolm, R.; Varrone, Z.; Collis, A.; Minto, M.; Burgess, S.; McDaniel, L.; Stimpson, E.; Spriggs, F.; Williams, J.; Neurath, K.; Ioime, N.; Agee, M.; Voss, E.; Furtak, K.; Renzulli, R.; Aanensen, N.; Carrolla, S.; Bickelhaupt, E.; Lazovatsky, Y.; DaSilva, A.; Zhong, J.; Stanyon, C. A.; Finley, R. L.; White, K. P.; Braverman, M.; Jarvie, T.; Gold, S.; Leach, M.; Knight, J.; Shimkets, R. A.; McKenna, M. P.; Chant, J.; Rothberg, J. M., A protein interaction map of Drosophila melanogaster. *Science* **2003,** *302* (5651), 1727-1736.
134. Skube, S. B.; Chaverri, J. M.; Goodson, H. V., Effect of GFP Tags on the Localization of EB1 and EB1 Fragments In Vivo. *Cytoskeleton* **2010,** *67* (1), 1-12.
135. Zhang, F. M.; Moniz, H. A.; Walcott, B.; Moremen, K. W.; Wang, L. C.; Linhardt, R. J., Probing the impact of GFP tagging on Robo1-heparin interaction. *Glycoconjugate Journal* **2014,** *31* (4), 299-307.
136. Griffin, B. A.; Adams, S. R.; Tsien, R. Y., Specific covalent labeling of recombinant protein molecules inside live cells. *Science* **1998,** *281* (5374), 269-272.
137. Hoffmann, C.; Gaietta, G.; Zuern, A.; Adams, S. R.; Terrillon, S.; Ellisman, M. H.; Tsien, R. Y.; Lohse, M. J., Fluorescent labeling of tetracysteine-tagged proteins in intact cells. *Nature Protocols* **2010,** *5* (10), 1666-1677.
138. Scheck, R. A.; Schepartz, A., Surveying Protein Structure and Function Using Bis-Arsenical Small Molecules. *Accounts of Chemical Research* **2011,** *44* (9), 654-665.
139. Stroffekova, K.; Proenza, C.; Beam, K. G., The protein-labeling reagent FLASH-EDT2 binds not only to CCXXCC motifs but also non-specifically to endogenous cysteine-rich proteins. *Pflugers Archiv-European Journal of Physiology* **2001,** *442* (6), 859-866.
140. Adams, S. R.; Campbell, R. E.; Gross, L. A.; Martin, B. R.; Walkup, G. K.; Yao, Y.; Llopis, J.; Tsien, R. Y., New biarsenical Ligands and tetracysteine motifs for protein labeling in vitro and in vivo: Synthesis and biological applications. *Journal of the American Chemical Society* **2002,** *124* (21), 6063-6076.
141. Van Engelenburg, S. B.; Nahreini, T.; Palmer, A. E., FACS-Based Selection of Tandem Tetracysteine Peptides with Improved ReAsH Brightness in Live Cells. *Chembiochem* **2010,** *11* (4), 489-493.
142. Martin, B. R.; Giepmans, B. N. G.; Adams, S. R.; Tsien, R. Y., Mammalian cell-based optimization of the biarsenical-binding tetracysteine motif for improved fluorescence and affinity. *Nature Biotechnology* **2005,** *23* (10), 1308-1314.
143. Soh, N., Selective chemical labeling of proteins with small fluorescent molecules based on metal-chelation methodology. *Sensors* **2008,** *8* (2), 1004-1024.

144. Ignatova, Z.; Gierasch, L. M., Monitoring protein stability and aggregation in vivo by real-time fluorescent labeling. *Proceedings of the National Academy of Sciences of the United States of America* **2004,** *101* (2), 523-528.
145. Taguchi, Y.; Shi, Z.-D.; Ruddy, B.; Dorward, D. W.; Greene, L.; Baron, G. S., Specific Biarsenical Labeling of Cell Surface Proteins Allows Fluorescent- and Biotin-tagging of Amyloid Precursor Protein and Prion Proteins. *Molecular Biology of the Cell* **2009,** *20* (1), 233-244.
146. Ray-Saha, S.; Schepartz, A., Visualizing Tyrosine Kinase Activity with Bipartite Tetracysteine Display. *Chembiochem* **2010,** *11* (15), 2089-2091.
147. Thorn, K. S.; Naber, N.; Matuska, M.; Vale, R. D.; Cooke, R., A novel method of affinity-purifying proteins using a bis-arsenical fluorescein. *Protein Science* **2000,** *9* (2), 213-217.
148. Zurn, A.; Klenk, C.; Zabel, U.; Reiner, S.; Lohse, M. J.; Hoffmann, C., Site-Specific, Orthogonal Labeling of Proteins in Intact Cells with Two Small Biarsenical Fluorophores. *Bioconjugate Chemistry* **2010,** *21* (5), 853-859.
149. Rutkowska, A.; Haering, C. H.; Schultz, C., A FlAsH-Based Cross-Linker to Study Protein Interactions in Living Cells. *Angewandte Chemie-International Edition* **2011,** *50* (52), 12655-12658.
150. Luedtke, N. W.; Dexter, R. J.; Fried, D. B.; Schepartz, A., Surveying polypeptide and protein domain conformation and association with FlAsH and ReAsH. *Nature Chemical Biology* **2007,** *3* (12), 779-784.
151. Doerr, A., Smaller can be better. *Nat Meth* **2008,** *5* (1), 6-7.
152. Goodman, J. L.; Fried, D. B.; Schepartz, A., Bipartite Tetracysteine Display Requires Site Flexibility for ReAsH Coordination. *Chembiochem* **2009,** *10* (10), 1644-1647.
153. Shaikh, T. A.; Bakus, R. C.; Parkin, S.; Atwood, D. A., Structural characteristics of 2-halo-1,3,2-dithiarsenic compounds and tris-(pentafluorophenylthio)-arsen. *Journal of Organometallic Chemistry* **2006,** *691* (9), 1825-1833.
154. Scheck, R. A.; Lowder, M. A.; Appelbaum, J. S.; Schepartz, A., Bipartite Tetracysteine Display Reveals Allosteric Control of Ligand-Specific EGFR Activation. *Acs Chemical Biology* **2012,** *7* (8), 1367-1376.
155. Avraham, R.; Yarden, Y., Feedback regulation of EGFR signalling: decision making by early and delayed loops. *Nature Reviews Molecular Cell Biology* **2011,** *12* (2), 104-117.
156. Lowder, M. A.; Appelbaum, J. S.; Hobert, E. M.; Schepartz, A., Visualizing protein partnerships in living cells and organisms. *Current Opinion in Chemical Biology* **2011,** *15* (6), 781-788.
157. Chen, I.; Ting, A. Y., Site-specific labeling of proteins with small molecules in live cells. *Current Opinion in Biotechnology* **2005,** *16* (1), 35-40.
158. Andresen, M.; Schmitz-Salue, R.; Jakobs, S., Short tetracysteine tags to beta-tubulin demonstrate the significance of small labels for live cell imaging. *Molecular Biology of the Cell* **2004,** *15* (12), 5616-5622.
159. Pomorski, A.; Otlewski, J.; Krezel, A., The High Zn-II Affinity of the Tetracysteine Tag Affects Its Fluorescent Labeling with Biarsenicals. *Chembiochem* **2010,** *11* (9), 1214-1218.
160. Enninga, J.; Mounier, J.; Sansonetti, P.; Van Nhieu, G. T., Secretion of type III effectors into host cells in real time. *Nature Methods* **2005,** *2* (12), 959-965.
161. Mota, L. J.; Holden, D. W., FlAsHlights on bacterial virulence proteins. *Nature Methods* **2005,** *2* (12), 898-899.
162. Romantsov, T.; Battle, A. R.; Hendel, J. L.; Martinac, B.; Wood, J. M., Protein Localization in Escherichia coli Cells: Comparison of the Cytoplasmic Membrane Proteins ProP, LacY, ProW, AqpZ, MscS, and MscL. *Journal of Bacteriology* **2010,** *192* (4), 912-924.

163. Romantsov, T.; Helbig, S.; Culham, D. E.; Gill, C.; Stalker, L.; Wood, J. M., Cardiolipin promotes polar localization of osmosensory transporter ProP in Escherichia coli. *Molecular Microbiology* **2007,** *64* (6), 1455-1465.
164. Xiong, Y.; Chen, B.; Shi, L.; Fredrickson, J. K.; Bigelow, D. J.; Squier, T. C., Targeted Protein Degradation of Outer Membrane Decaheme Cytochrome MtrC Metal Reductase in Shewanella oneidensis MR-1 Measured Using Biarsenical Probe CrAsH-EDT2. *Biochemistry* **2011,** *50* (45), 9738-9751.
165. List, B.; Barbas, C. F.; Lerner, R. A., Aldol sensors for the rapid generation of tunable fluorescence by antibody catalysis. *Proceedings of the National Academy of Sciences of the United States of America* **1998,** *95* (26), 15351-15355.
166. Walter, K. U.; Vamvaca, K.; Hilvert, D., An active enzyme constructed from a 9-amino acid alphabet. *Journal of Biological Chemistry* **2005,** *280* (45), 37742-37746.
167. Gibson, D. G.; Glass, J. I.; Lartigue, C.; Noskov, V. N.; Chuang, R.-Y.; Algire, M. A.; Benders, G. A.; Montague, M. G.; Ma, L.; Moodie, M. M.; Merryman, C.; Vashee, S.; Krishnakumar, R.; Assad-Garcia, N.; Andrews-Pfannkoch, C.; Denisova, E. A.; Young, L.; Qi, Z.-Q.; Segall-Shapiro, T. H.; Calvey, C. H.; Parmar, P. P.; Hutchison, C. A.; Smith, H. O.; Venter, J. C., Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science* **2010,** *329* (5987), 52-56.
168. Appella, D. H.; Christianson, L. A.; Karle, I. L.; Powell, D. R.; Gellman, S. H., β-Peptide Foldamers: Robust Helix Formation in a New Family of β-Amino Acid Oligomers. *Journal of the American Chemical Society* **1996,** *118* (51), 13071-13072.
169. Ho, S. P.; DeGrado, W. F., Design of a 4-helix bundle protein: synthesis of peptides which self-associate into a helical protein. *Journal of the American Chemical Society* **1987,** *109* (22), 6751-6758.
170. Regan, L.; Degrado, W. F., Characterization of a helical protein designed from 1st principles. *Science* **1988,** *241* (4868), 976-978.
171. Farid, T. A.; Kodali, G.; Solomon, L. A.; Lichtenstein, B. R.; Sheehan, M. M.; Fry, B. A.; Bialas, C.; Ennist, N. M.; Siedlecki, J. A.; Zhao, Z.; Stetz, M. A.; Valentine, K. G.; Anderson, J. L. R.; Wand, A. J.; Discher, B. M.; Moser, C. C.; Dutton, P. L., Elementary tetrahelical protein design for diverse oxidoreductase functions. *Nat Chem Biol* **2013,** *9* (12), 826-833.
172. Koder, R. L.; Anderson, J. L. R.; Solomon, L. A.; Reddy, K. S.; Moser, C. C.; Dutton, P. L., Design and engineering of an O2 transport protein. *Nature* **2009,** *458* (7236), 305-309.
173. Zaccai, N. R.; Chi, B.; Thomson, A. R.; Boyle, A. L.; Bartlett, G. J.; Bruning, M.; Linden, N.; Sessions, R. B.; Booth, P. J.; Brady, R. L.; Woolfson, D. N., A de novo peptide hexamer with a mutable channel. *Nature Chemical Biology* **2011,** *7* (12), 935-941.
174. Fletcher, J. M.; Harniman, R. L.; Barnes, F. R. H.; Boyle, A. L.; Collins, A.; Mantell, J.; Sharp, T. H.; Antognozzi, M.; Booth, P. J.; Linden, N.; Miles, M. J.; Sessions, R. B.; Verkade, P.; Woolfson, D. N., Self-Assembling Cages from Coiled-Coil Peptide Modules. *Science* **2013,** *340* (6132), 595-599.
175. Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Röthlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Barbas, C. F.; Hilvert, D.; Houk, K. N.; Stoddard, B. L.; Baker, D., De Novo Computational Design of Retro-Aldol Enzymes. *Science* **2008,** *319* (5868), 1387-1391.
176. Rothlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S.; Baker, D., Kemp elimination catalysts by computational enzyme design. *Nature* **2008,** *453* (7192), 190-195.
177. Fleishman, S. J.; Whitehead, T. A.; Ekiert, D. C.; Dreyfus, C.; Corn, J. E.; Strauch, E.-M.; Wilson, I. A.; Baker, D., Computational Design of Proteins Targeting the Conserved Stem Region of Influenza Hemagglutinin. *Science* **2011,** *332* (6031), 816-821.

178. King, N. P.; Sheffler, W.; Sawaya, M. R.; Vollmar, B. S.; Sumida, J. P.; André, I.; Gonen, T.; Yeates, T. O.; Baker, D., Computational Design of Self-Assembling Protein Nanomaterials with Atomic Level Accuracy. *Science* **2012,** *336* (6085), 1171-1174.
179. Leung, D. W., E. Chen, and D. V. Goeddel., A method for the random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique* **1989,** (1), 11-15.
180. Eckert, K. A.; Kunkel, T. A., DNA polymerase fidelity and the polymerase chain reaction. *PCR methods and applications* **1991,** *1* (1), 17-24.
181. Cadwell, R. C.; Joyce, G. F., Randomization of genes by PCR mutagenesis. *PCR methods and applications* **1992,** *2* (1), 28-33.
182. McCullum, E.; Williams, B. R.; Zhang, J.; Chaput, J., Random Mutagenesis by Error-Prone PCR. In *In Vitro Mutagenesis Protocols*, Braman, J., Ed. Humana Press: 2010; Vol. 634, pp 103-109.
183. Hogrefe, H., Fine-Tuning Enzyme Activity Through Saturation Mutagenesis. In *In Vitro Mutagenesis Protocols*, Braman, J., Ed. Humana Press: 2010; Vol. 634, pp 271-283.
184. Wang, D.; Zhao, C.; Cheng, R.; Sun, F. Z., Estimation of the mutation rate during error-prone polymerase chain reaction. *Journal of Computational Biology* **2000,** *7* (1-2), 143-158.
185. Chen, K. Q.; Arnold, F. H., Tuning the activity of an enzyme for unusual environments - sequential random mutagenesis of Subtilisin-E for catalysis in dimethylformamide. *Proceedings of the National Academy of Sciences of the United States of America* **1993,** *90* (12), 5618-5622.
186. You, L.; Arnold, F. H., Directed evolution of subtilisin E in Bacillus subtilis to enhance total activity in aqueous dimethylformamide. *Protein Engineering* **1996,** *9* (1), 77-83.
187. Moore, J. C.; Arnold, F. H., Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents. *Nature Biotechnology* **1996,** *14* (4), 458-467.
188. Farinas, Edgardo T.; Schwaneberg, U.; Glieder, A.; Arnold, Frances H., Directed Evolution of a Cytochrome P450 Monooxygenase for Alkane Oxidation. *Advanced Synthesis & Catalysis* **2001,** *343* (6-7), 601-606.
189. Glieder, A.; Farinas, E. T.; Arnold, F. H., Laboratory evolution of a soluble, self-sufficient, highly active alkane hydroxylase. *Nature Biotechnology* **2002,** *20* (11), 1135-1139.
190. Stemmer, W. P. C., Rapid evolution of a protein in-vitro by dna shuffling. *Nature* **1994,** *370* (6488), 389-391.
191. Erijman, A.; Dantes, A.; Bernheim, R.; Shifman, J. M.; Peleg, Y., Transfer-PCR (TPCR): A highway for DNA cloning and protein engineering. *Journal of Structural Biology* **2011,** *175* (2), 171-177.
192. Erijman, A.; Shifman, J. M.; Peleg, Y., A single-tube assembly of DNA using the transfer-PCR (TPCR) platform. *Methods in molecular biology (Clifton, N.J.)* **2014,** *1116*, 89-101.
193. Singh, M. I.; Jain, V., Tagging the Expressed Protein with 6 Histidines: Rapid Cloning of an Amplicon with Three Options. *Plos One* **2013,** *8* (5).
194. Wan, X. B.; Zhang, W.; Li, L.; Xie, Y. T.; Li, W.; Huang, N., A New Target for an Old Drug: Identifying Mitoxantrone as a Nanomolar Inhibitor of PIM1 Kinase via Kinome-Wide Selectivity Modeling. *Journal of Medicinal Chemistry* **2013,** *56* (6), 2619-2629.
195. Aizner, Y.; Sharabi, O.; Shirian, J.; Dakwar, G. R.; Risman, M.; Avraham, O.; Shifman, J., Mapping of the Binding Landscape for a Picomolar Protein-Protein Complex through Computation and Experiment. *Structure* **2014,** *22* (4), 636-645.
196. Dahiyat, B. I.; Mayo, S. L., De Novo Protein Design: Fully Automated Sequence Selection. *Science* **1997,** *278* (5335), 82-87.

197. Yosef, E.; Politi, R.; Choi, M. H.; Shifman, J. M., Computational Design of Calmodulin Mutants with up to 900-Fold Increase in Binding Specificity. *Journal of Molecular Biology* **2009,** *385* (5), 1470-1480.
198. Hayashi, N.; Welschof, M.; Zewe, M.; Braunagel, M.; Dubel, S.; Breitling, F.; Little, M., Simultaneous mutagenesis of antibody CDR regions by overlap extension and PCR. *Biotechniques* **1994,** *17* (2), 310-&.
199. Miyazaki, K.; Arnold, F. H., Exploring nonnatural evolutionary pathways by saturation mutagenesis: Rapid improvement of protein function. *Journal of Molecular Evolution* **1999,** *49* (6), 716-720.
200. Hibbert, E. G.; Baganz, F.; Hailes, H. C.; Ward, J. M.; Lye, G. J.; Woodley, J. M.; Dalby, P. A., Directed evolution of biocatalytic processes. *Biomolecular Engineering* **2005,** *22* (1-3), 11-19.
201. Hibbert, E. G.; Senussi, T.; Costelloe, S. J.; Lei, W.; Smith, M. E. B.; Ward, J. M.; Hailes, H. C.; Dalby, P. A., Directed evolution of transketolase activity on non-phosphorylated substrates. *Journal of Biotechnology* **2007,** *131* (4), 425-432.
202. Hibbert, E. G.; Senussi, T.; Smith, M. E. B.; Costelloe, S. J.; Ward, J. M.; Hailes, H. C.; Dalby, P. A., Directed evolution of transketolase substrate specificity towards an aliphatic aldehyde. *Journal of Biotechnology* **2008,** *134* (3-4), 240-245.
203. Smith, M. E. B.; Hibbert, E. G.; Jones, A. B.; Dalby, P. A.; Hailes, H. C., Enhancing and Reversing the Stereoselectivity of Escherichia coli Transketolase via Single-Point Mutations. *Advanced Synthesis & Catalysis* **2008,** *350* (16), 2631-2638.

Chapter 2

Application of ferritins to nanotechnology

NOTE: Parts of this chapter were undertaken with the help of collaborators Prof. Phil Blower and Dr. Maggie Cooper (KCL). All except the mouse injection, SPECT scan/reconstruction and culling of the mice were untaken by this PhD candidate.

2.1. Introduction

2.1.1. Aims of this investigation

Designing and creating new tools to answer scientific questions or to solve engineering problems in new ways is limited by the technology currently available. Nature, over the course of evolution has converged on processes and tools to complete a wide range of tasks sometimes in a more effective or efficient manner than what is possible with man-made processes. Therefore one strategy to enable science could be to learn and adopt methods from Nature and adapt them to fulfill current needs.

This research described in this chapter aims to expand the utility of protein nanocages which, as discussed in Chapter 1, have already been applied to a number of applications. Previously, The Orner laboratory demonstrated that wild type nanocages, without any modifications to their primary sequence, could be used to create monodispersed protein encapsulated gold nanoparticles[1]. While this work pushed forward the boundaries of protein nanoparticle research, this method had not been expanded outside of the single protein example. Work described in this chapter aims to broaden this method to a larger repertoire of nanocages. If successful, it would demonstrate the breath and robustness of the technique. If not successful, it would hopefully shed light on and provide hints on overcoming some of the pitfalls and challenges with applying these proteins to the production of nanomaterials.

The second aim of this investigation was to perform early stage explorations into the potential of using nanocages for biomedical applications. It has been proposed that hollow protein nanocages possess the potential for drug delivery applications. In addition, a long term use of the protein encapsulated nanoparticles described above could be *in vivo* imaging. In

both technologies, it would be useful to direct these proteins to specific regions of the body most likely through the use of fusions to tissue-specific ligands. However, as an initial step to these ends, an understanding of the native preference of protein nanocages for the various bodily tissues must first be established, and this is the second aim of this project.

2.1.2. Ferritin nanocage proteins: Horse spleen ferritin (HsFn), Bacterioferritin (Bfr) and DNA binding protein from starved cells (Dps)

As discussed in Chapter 1, the ferritin superfamily provides excellent models for the study of quaternary structure due to the relativity simple fold of their monomers. Most ferritins assemble into nano-scaled, spherical nanocage structures which play a role in maintaining cellular iron homeostasis[2, 3]. The three ferritins described below are the main proteins studied and utilized in this thesis.

In mammalian cells, ferritins are heterooligomers assembled from both a light (L) and heavy (H) chain, where the light chain is used as a structural scaffold and the heavy chain contains a ferroxidase catalytic centre where iron is oxidized from Fe^{2+} to Fe^{3+}. Most mammalian ferritin is localized inside the cell, however, some is excreted to the serum, and serum ferritin levels can be used to approximate iron concentration of the body[4, 5]. Mammalian ferritin monomers fold into a four-helix bundle, like all ferritins, and have a short, C-terminal helix (E-helix) positioned perpendicular to the main body of the bundle. Twenty-four of these monomers assemble into an octahedral maxi-ferritin nanocage[6]. A widely studied mammalian ferritin, and one of the earliest examples to be isolated and characterised, is horse spleen ferritin (HsFn, see Appendix 2.1 for full sequence)[7, 8]. Although this protein assembles in the cell from a mixture of L- and H-chain monomers each can form a homomeric nanocage alone. Although the L-chain nanocage is non-catalytic, it is highly stable[9]. The highly stable and readily available HsFn L-chain nanocage has been a pillar of nanocage research[1, 10].

Ferritins from bacteria, in contrast to mammalian ferritins, form homomeric nanocages. Bacteria have maxi-ferritins that assemble into heme- and non-heme-containing nanocages. This thesis will primarily use a heme-containing ferritin, or bacterioferritin, from *E. coli* (Bfr, see Appendix 2.1 for full sequence) [11, 12]. The Bfr monomer folds into a four-helix bundle with a C-terminal fifth, E-helix, running 78 ^{0} to the main bundle axis. The E-helix points into the nanocage at the four-fold symmetry axis. Each monomer contains one ferroxidase active site[13, 14, 15], and once assembled into a 12 nm octahedral nanocage, the resulting 9 nm inner cavity can store up to 4500 mineralized iron atoms. Analysis of Bfr with size exclusion chromatography (SEC) shows that it exists in solution as a mixture of nanocage and a dimer oligomerization states, and it is believed that the nanocage assembles through the association of two-fold symmetric dimers.

The mini-ferritin, DNA binding protein from starved cells (Dps, see Appendix 2.1 for sequence) is another ferritin present in *E.coli*[16, 17, 18]. This protein assembles from twelve identical monomers into a 9 nm nanocage with tetrahedral symmetry and a 4.5 nm internal cavity. The Dps monomer folds into a four helix bundle with remarkable similarity to the tertiary fold of the Bfr monomer, however, Dps has a short additional helix (BC-helix) between the B- and C-helices and lacks the E-helix of Bfr. The BC-helix is located on the outside of the assembled nanocage at the two-fold symmetry axis and BC-helices from two monomers form part of the dimer protein-protein interface[19]. Unlike Bfr, Dps is a single oligomerization state in solution and although the dimer species is not observable, is it believed that, like Bfr, Dps also assembles from dimer subunits. While this nanocage is also capable of storing mineralized iron, the protein is only expressed as a response to oxidative stress. Interestingly, Dps has been shown to bind DNA nonspecifically and help protect it from reactive oxygen species[20, 21, 22].

2.1.3. Formation of inorganic nanoparticles inside ferritins

The size and shape of an inorganic nanoparticle greatly affects its properties and, thus, much research has focused on trying to control these properties as discussed in Chapter 1.

Nanocage proteins, because of their defined size and shape, have been used as size-constrained reactors to afford nanoparticles with very narrow polydispersities. In addition, they provide the resulting encapsulated particles with enhanced solubility in biological media, protection from aggregation, and enable the further manipulation or purification of the particles through bioconjugate and fusion chemistries. Ferritins, as discussed above and in Chapter 1, are examples of well-defined nanocages and have been used for nanoparticle research. A wide range of nanoscaled materials have been generated inside ferritins. Palladium[23] and cadmium selenide[24] particles have been created by exploiting ferritin affinity for metal cations, and gold particles have been generated through reagent photo-activation in the presence of the mineralized iron core[25]. However, these methodologies have been shown to be non-specific in that nanoparticles can form both inside and outside of the nanocage. Dmochowski and co-workers[26] described conditions that resulted in polydisperse gold nanostructures forming on the outside of the ferritin nanocage and were only able to achieve particle formation inside though engineering the protein to remove metal binding residues such as cysteine or histidine on the outside and to position additional copies of these residues inside of the nanocage[27]. They also produced ferritin encapsulated gold nanoparticles by assembling ferritin subunits around pre-generated particles[28]. Naik and co-workers grew silver nanoparticle selectivity inside ferritins by fusing a silver binding peptide to the terminus which is located at the inner cavity of the ferritin[29].

These latter methods, while successful, have the drawback that protein engineering is required to elicit specificity inside the ferritin nanocage. A method with the most utility would be one that can generate nanoparticles in the inner cavity of any nanocage without any engineering required. The Orner laboratory has developed a method to create gold nanoparticles inside unmodified horse spleen light chain ferritins with the intention that it would be general for any protein nanocage[1]. This method initially used the HsFn L-chain nanocage which is especially stable, non-catalytically active and assembles into a single

oligomerization state in solution. The first step in the synthetic strategy involves the slow diffusion of gold ions into the ferritin cavity. Because the diffusion is slow, the protein solution can be rapidly desalted to remove gold external to the nanocage. Then the trapped gold is reduced with a fast-reducing agent, trapping it as a nanocluster inside the nanocage. More gold feedstock is added and a weaker reducing agent is then added to allow the nanocluster to grow slowly into a nanoparticle the size of the inner cavity. This method generated ferritin encapsulated gold nanoparticles with narrow polydispersities with little damage to the protein (Figure 2.1).

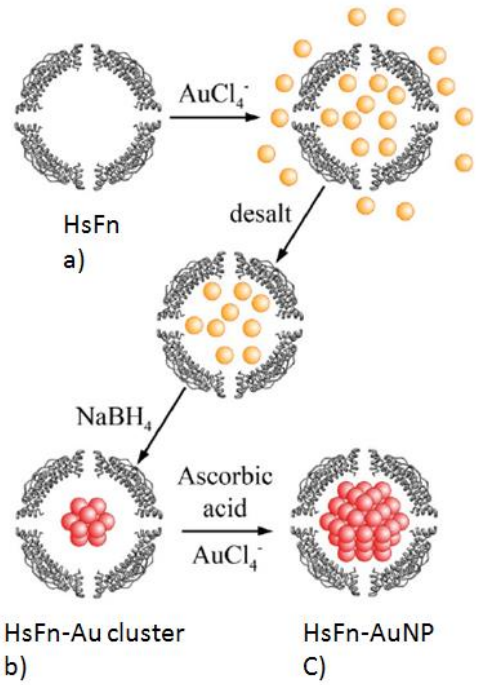


Figure 2.1 – Schematic of the two-step reduction process to create monodisperse gold nanoparticles inside HsFn. (a) Incubation of $HAuCl_4$ with the protein, followed by desalting of the protein sample to remove any external gold atoms. The remaining gold is located inside the ferritin nanocage which is reduced into a nanocluster using the first reducing agent, $NaBH_4$. (b) Addition of extra $HAuCl_4$ and the second reducing agent, ascorbic acid, to grow the nanocluster into a nanoparticle. (c) The formation of monodisperse gold nanoparticles inside of HsFn. This figure is adapted from figures in Fan et al (2010)[1].

This method is simple in its execution and could be a powerful tool for the generation of protein-encapsulated nanoparticles and monodisperse nanoparticles (Figure 2.2), however, to date, while a similar method has been used with copper,[30] the generality of the approach has not been definitively established in that it has been limited to only HsFn. Therefore, to

increase this methods utility to the scientific community, it should be optimised for nanocages of different sizes and made up of different proteins. This chapter will describe attempts to achieve these goals.

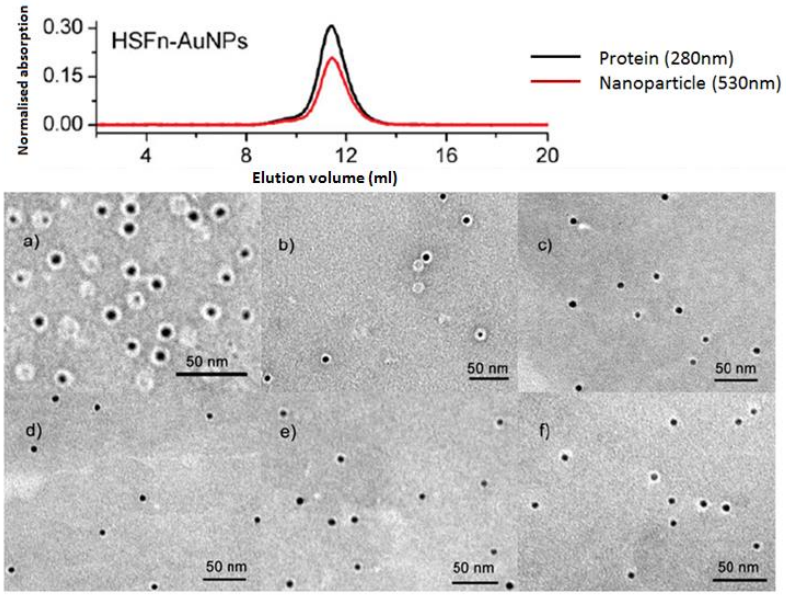


Figure 2.2 – Mono-dispersed gold nanoparticles generated and encapsulated inside HsFn protein nanocages. (top) SEC analysis showing the co-elution of protein nanocages (black) and nanoparticles (red). (bottom) Negative stained TEM analysis of HsFn encapsulated gold nanoparticles, showing the protein nanocages as pale rings and the gold nanoparticles electron rich black circles. Encapsulated nanoparticles have white rings surrounding the black circles. (a – b) before sucrose gradient purification, (c – f) after sucrose gradient purification, selecting for nanoparticles. This figure is adapted from figures in Fan et al (2010)[1] and its supporting information.

2.1.4. Functionalisation of nanocages for *in vivo* visualization of localisation

A key attribute of protein nanocages that makes them useful for many application is their multiple surface character; they have an external and internal surface, that often can be modified independently from each other[10]. Thus their characteristics, and resulting functions, can be expanded in sometimes orthogonal ways.

To date, this feature of ferritins has allowed them to fulfil a range of *in vivo* tasks in both medical imaging and therapy. Ferritins encapsulating ferromagnetic iron oxide nanoparticles have been used as magnetic resonance imaging (MRI) contrasting agents[31]. Recent studies combined these nanoparticles, with a surface fusion of the green fluorescent

protein (GFP) and the cancer targeting peptide RGD, to create a tumour cell targeting MRI and fluorescent imaging probe[32] similar to work reported using a multifunctional liposome[33].

The inner cavities of ferritins have also been used for radiolabelling. Radioactive Cu^{64} was non selectivity loaded into the inner cavity of a protein nanocage which had the cancer cell targeting peptide RGD fused to its surface, with its uptake monitored over time[34]. Similarly, a RGD fusion human H chain ferritin containing a ferromagnetic iron oxide nanoparticle was used to monitor the targeting of cancer cells[35], inflammation and angiogenesis[36].

The above examples demonstrate the potential of ferritins for medical imaging; however they all require filling the nanocage with the imaging agent. This can be a drawback, especially with the realisation that a labelled ferritin could have in concert multiple functions of imaging, delivery, and direction. Therefore we thought that a better strategy would be to attach the label to the outside of the nanocage so that the inside would be preserved for eventual delivery and transport applications. This chapter will describe a first step toward the development of this technology where we will establish the labelling and imaging and determine any native targeting of a ferritin which presents no engineered tissue directing fusions.

2.2. Results and discussions

2.2.1. Generation of gold nanoparticles inside Bfr and Dps

Our laboratory has extensive experience working with the protein nanocages Bfr and Dps. This experience and the fact that one is a maxi-ferritin and the other is a mini-ferritin led us to choose them as initial model systems to expand our method for generating encapsulated gold nanoparticles of different sizes. In addition, it should be recalled that the ferritin that was previously used was HsFn L-chain which has no catalytic activity. Both Bfr and Dps, however, are catalytic and can bind metal. Furthermore, Bfr contains multiple oligomerization states. [37, 38] We thought these attributes of the two proteins would present a challenge to test the robustness of our methodology.

Our analytical techniques were similar to those that were used previously in the first generation of this synthetic method. Size exclusion chromatography (SEC) was used as the primary analytical tool. First, SEC can detect the presence of the protein (280 nm), and through separation, identify the size and number of species present (aggregate, nanocage, dimer, monomer, etc). Second surface plasmon resonance (SPR) can be used to detect the nanoparticles (530 nm). Overlaying the SEC traces from the two channels can show which state of protein the gold is associated. The secondary analytical tool was transmission electron microscopy (TEM) to determine morphology of gold structures and the state of the protein. First, standard elution volumes of intact Bfr and Dps needed to be established.

The proteins Bfr and Dps were expressed, purified and characterised (see Appendix 2.2 and 2.3 for SDS-PAGE and CD respectively). SEC analysis of Bfr showed, consistent with the literature, that two species are present[37]. Both the 24-mer nanocage (elution volume 11 ml) and a dimer species (elution volume 15.5 ml) are observed, while for Dps only a single 12-mer nanocage species (elution volume13 ml) exists in solution (Figure 2.3). These elution profiles for Bfr and Dps nanocages were used as standards for the research described in this section.

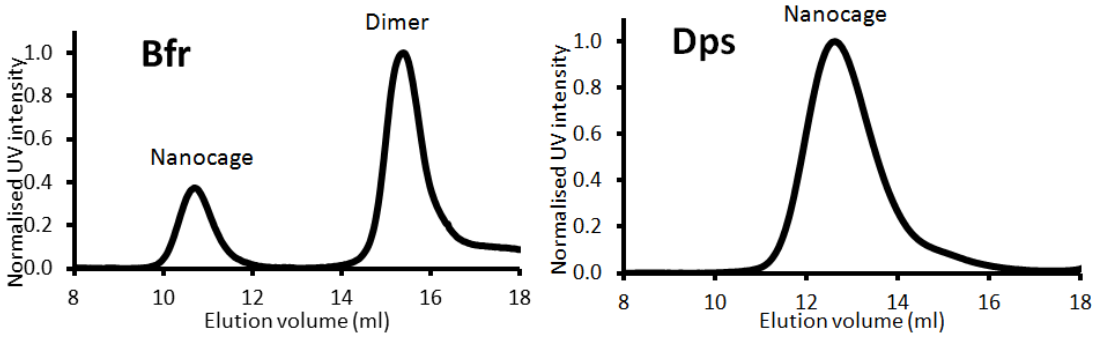


Figure 2.3 – Size exclusion chromatograms of purified Bfr (left) and Dps (right). These chromatograms are standards that were used to assess the survival of the proteins under nanoparticle generation conditions. The chromatograms were normalised to their highest intensity. The data is the average of three runs (0.5 ml injection of 1 mg/ml protein(280 nm), using running buffer, 50 mM Tris.HCl, 50 mM NaCl, pH 7.8, at a flow rate of 0.5 ml/min.

We initially employed the unoptimised HsFn conditions using Bfr and Dps. However, SEC showed only large protein aggregates for both proteins with co-eluting SPR peaks (Figure 2.4) and no peaks from surviving nanocages or dimers were observed. This surprising result immediately highlighted the difference in behaviour between HsFn and the bacterial ferritins. A small aggregate peak can be observed when HsFn is subjected to these conditions, however it is much smaller than with Bfr and Dps and nanocage is the major peak[1]. Because of this initial failure, the components of the conditions were optimised.

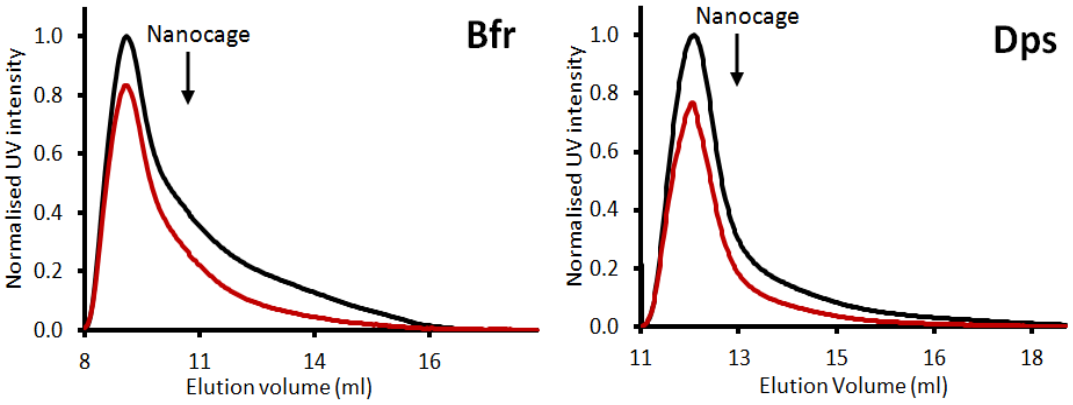


Figure 2.4 – Initial gold nanoparticle genesis conditions cause protein aggregation. Size exclusion chromatograms of purified Bfr (left) and Dps (right) after being subjected to nanoparticle generation conditions. The chromatograms normalised to their highest intensity. 280 nm (black) and 530 nm (red). Nanoparticle experiment from trial 1, used 1 ml of 1mg/ml protein in phosphate buffer (50 mM Na_2HPO_4, 50 mM NaCl, pH 7) incubated with 22.6 µl of 0.1 M $HAuCl_4$ for 3 hrs, at room temp. 20 µl of 0.1 M $NaBH_4$ was then added and was incubated for 3 hrs, at room temp followed by desalting. An additional 10 µl of 0.1 M $HAuCl_4$ and 30 µl of 0.1 M ascorbic acid was added and incubated, overnight at room temp. All reagent solutions are in H_2O. After concentration, 0.5 ml of the remaining sample was injected. Each sample was run once.

2.2.1.1. Optimisation of nanoparticle generation conditions for protein stability

A number of changes were made in the initial conditions with the hope of trying to eliminate protein aggregation. There are several areas for optimisation (Figure 2.5), including the reagents, the reagent concentrations, volumes and buffering conditions. Each condition that was attempted was analysed with SEC for the presence of nanocages or protein aggregates. If nanocage peaks were observed the sample was further characterised with TEM (all conditions that were attempted are listed in Appendix 2.7).

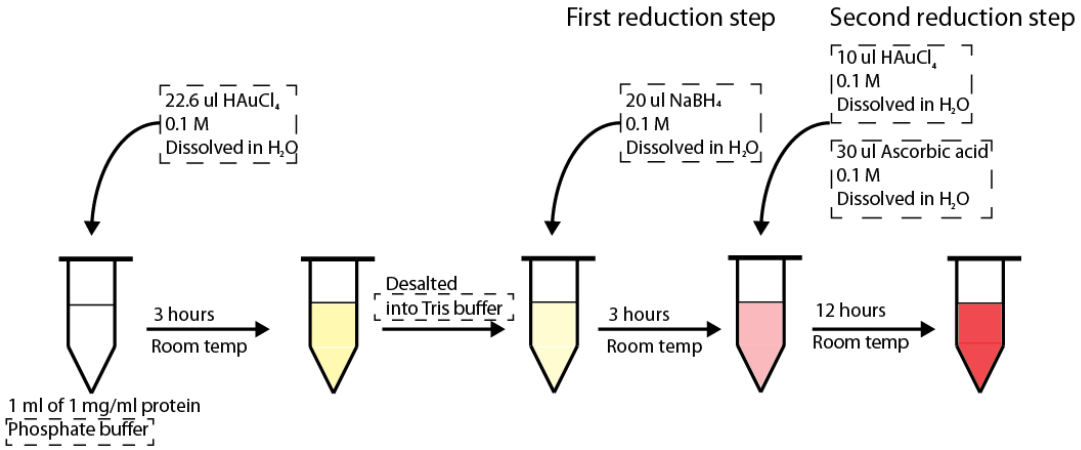


Figure 2.5 – Schematic representation of the experimental procedure to create gold nanoparticles inside of HsFn as previously reported[1] highlighting areas of possible optimisation (black dashed boxes). $HAuCl_4$ is first added to the protein sample, which is subsequently desalted and reduced with the addition of the first reducing agent. More $HAuCl_4$ is added and further reduced with the addition of the second reducing agent. All concentrations and volumes of the reagents are the added quantities and not the final. All the possible areas of optimisation are highlighted (black dashed boxes), including the concentration, volume and buffering condition of each reagent, plus the reagents themselves, as well as the protein buffer and desalting running buffer.

The first two conditions that were screened involved attempts to prevent aggregation. In the first set of conditions (Trial 1), the same method as previously published[1] was attempted at 4 °C however this resulted in protein aggregation as was seen above. In a second set of conditions (Trial 2) a detergent (0.1% tween20) was included, but again, only large protein aggregates were observed. Although potentially a quick fix, stopping the aggregation of the Bfr and Dps through either lowering the temperature or through detergent did not work.

Next we tried to determine if aggregation was caused by any specific reagent that was used for the generation of the nanoparticles. First the role of the first reducing agent, $NaBH_4$, in forming aggregates was examined. In the next group of conditions, the concentration of $NaBH_4$ was decreased from 0.1 M to 0.01 M (Trial 3) and 0.001 M (Trial 4). Again, only a large protein aggregate peak was observed. Next, the $HAuCl_4$ concentration was decreased from 0.1 M to 0.05 M $HAuCl_4$ (Trial 5), however, again, both proteins aggregated. Taken together, these trials suggested that at least one, if not all, of these reagents were not compatible with Bfr and Dps.

In the initial conditions, the reagents were dissolved in water before addition to the buffered protein solution. The gold source in our method is chloroauric acid ($HAuCl_4$). The pH of the 0.1 M $HAuCl_4$ solution in water is 1.2. We considered that the addition of a strongly acidic solution to the protein sample, although buffered, could cause a sudden or permanent drop in pH which could lead to protein aggregation. Potentially, by buffering the $HAuCl_4$ solution to a neutral pH, protein aggregation could be avoided. However, it was realised that changing the pH could affect the reduction potential of the reductant. Therefore a mini-screen was performed where the ability to reduce $HAuCl_4$ was determined in a number of conditions in the absence of protein.

For the screen, reagents dissolved in various buffers were assessed for their ability to cause a solution of $HAuCl_4$ to turn red in twenty-four hours, indicating reduction of $HAuCl_4$ (see Appendix 2.5). An assay like this was ideal because of its simple and rapid detectability and the small volumes that were used. The reductant, $NaBH_4$ was assessed at different concentrations in H_2O (unbuffered), sodium phosphate buffer (pH 7) and Tris buffer (pH 7.8) Phosphate buffered reductant, especially at lower concentrations, only poorly reduced gold whereas the reagent was successful in both Tris and in H_2O across all concentrations. Because of its buffering ability, Tris was used for further experiments.

Using Tris buffered solutions (pH 7.8) optimisation of $HAuCl_4$ was again attempted in the presence of the proteins. However, for conditions using buffered $HAuCl_4$ at both 0.1 M (Trial 6) and 0.05 M (Trial 7), again, only aggregation for both Bfr and Dps was observed.

All the optimisations of the method thus far only analysed the product at the end of the process. Thus, it was thought that to better understand the cause of the aggregation it would be better to check protein stability after the addition of each reagent. Two different concentrations of the Tris buffered $HAuCl_4$ solution (0.1 M and 0.05 M Trials 8 and 9 respectively) were added to the Bfr and Dps solutions. After a three hour incubation period, they were analysed by SEC. The resulting chromatograms indicated the presence of nanocages and dimers and very little aggregates for both proteins. This result clearly demonstrated that the buffered gold solution was not responsible for protein aggregation. A similar experiment was performed using $NaBH_4$ (0.1 M and 0.01 M Trial 10 and 11 respectively) however, after three hours of incubation, only protein aggregates were observed for Bfr. Chromatograms of the Dps solution, however, showed both nanocages and aggregates.

As a goal of this research is to find conditions that can be generally applied to any protein nanocage, it was important that we used reagents that permitted the survivability of both Bfr and Dps. An alternative reducing agent to $NaBH_4$ is $NaCNBH_3$. Due to the electron withdrawing properties of the cyanide group, $NaCNBH_3$ is less reactive than $NaBH_4$. This drop in reactivity could provide milder conditions that would be general for both Bfr and Dps. Therefore, $NaCNBH_3$ was screened for its ability to reduce gold in buffered and unbuffered conditions (see Appendix 2.6). The Tris buffer conditions were superior.

Analysis of the survivability of Bfr and Dps were conducted using $NaCNBH_3$ as the first reducing agent. The concentration of $NaCNBH_3$ was examined at 0.1 M (Trials 12, 14 and 15) and 0.01 M (Trial 13) as well as a change in volume, using a 20 µl addition (Trials 12 and 13) a 100 µl addition (Trial 14) and a 200 µl addition (Trial 15). After 3 hours of incubation,

nanocages could be seen for both Bfr and Dps. This highlights the success in using $NaCNBH_3$ as the first reducing agent to combat protein aggregation. Taking these data into account, a new experiment was preformed which buffered both the gold solutions and the new first reducing agent $NaCNBH_4$ (Trial 16). The SEC data showed a continual signal across a wide elution volume for both Dps and Bfr when monitoring at 280 nm, but with no clear nanocage peak. While an SPR signal peak could be observed, this did not elute at the volume expected for nanocages (Figure 2.6).

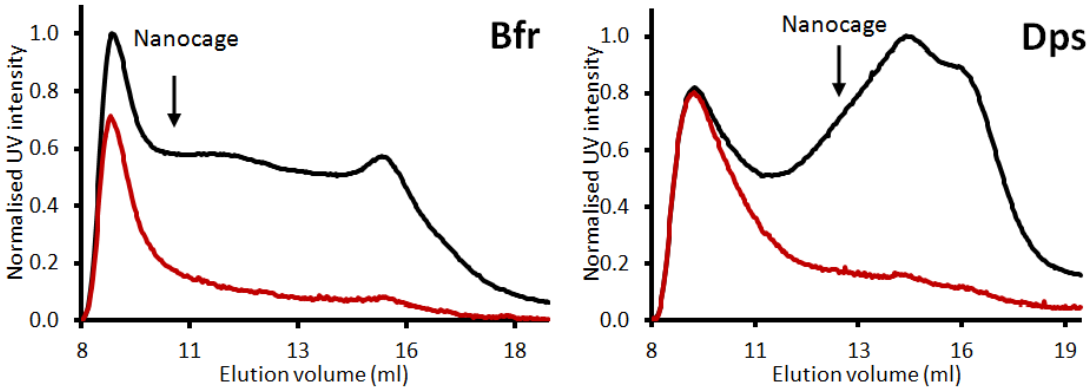


Figure 2.6 – Semi optimised gold nanoparticle genesis conditions buffering both the gold source and first reducing agent in Tris buffer (50 mM Tris, pH 7.8.). Size exclusion chromatograms of purified Bfr (left) and Dps (right) after being subjected to nanoparticle generation conditions. The chromatograms normalised to their highest intensity. 280 nm (black) and 530 nm (red). Nanoparticle experiment from trial 16, used 1 ml of 1mg/ml protein in Tris buffer (50 mM Tris.HCl, 50 mM NaCl, pH 7) incubated with 20 µl of 0.1 M $HAuCl_4$ buffered in Tris (50 mM Tris.HCl, pH7.8) for 3 hrs, at room temp. 20 µl of 0.1 M $NaCNBH_3$ buffered in Tris (50 mM Tris.HCl, pH7.8) was then added and was incubated for 3 hrs, at room temp followed by desalting. An additional 10 µl of 0.1 M $HAuCl_4$ buffered in Tris (50 mM Tris.HCl, pH7.8) and 30 µl of 0.1 M ascorbic acid was added and incubated, overnight at room temp. After concentration, 0.5 ml of the remaining sample was injected. Each sample was run once.

The lack of well-defined nanocage peaks in figure 2.6, made it unclear if the nanocages where aggregated or not. Samples were taken from the elution volume corresponding to the nanocages of both Bfr and Dps and analysed with TEM. Micrographs from TEM did not show the presence of any intact nanocages but did show nanoparticles with pale rings around them (Figure 2.7). These nanoparticles were of a fairly regular size and could be surrounded by folded nanocages. However the lack of empty but intact nanocages led us to believe that these nanoparticles were covered with protein aggregates. The similar size of the observed nanoparticles could be from the size selective nature of the purification process that was used

to isolate them. Only conditions that allowed for the observance of both empty and filled white rings (nanocages) would be seen as optimal and proof of nanocage survival.

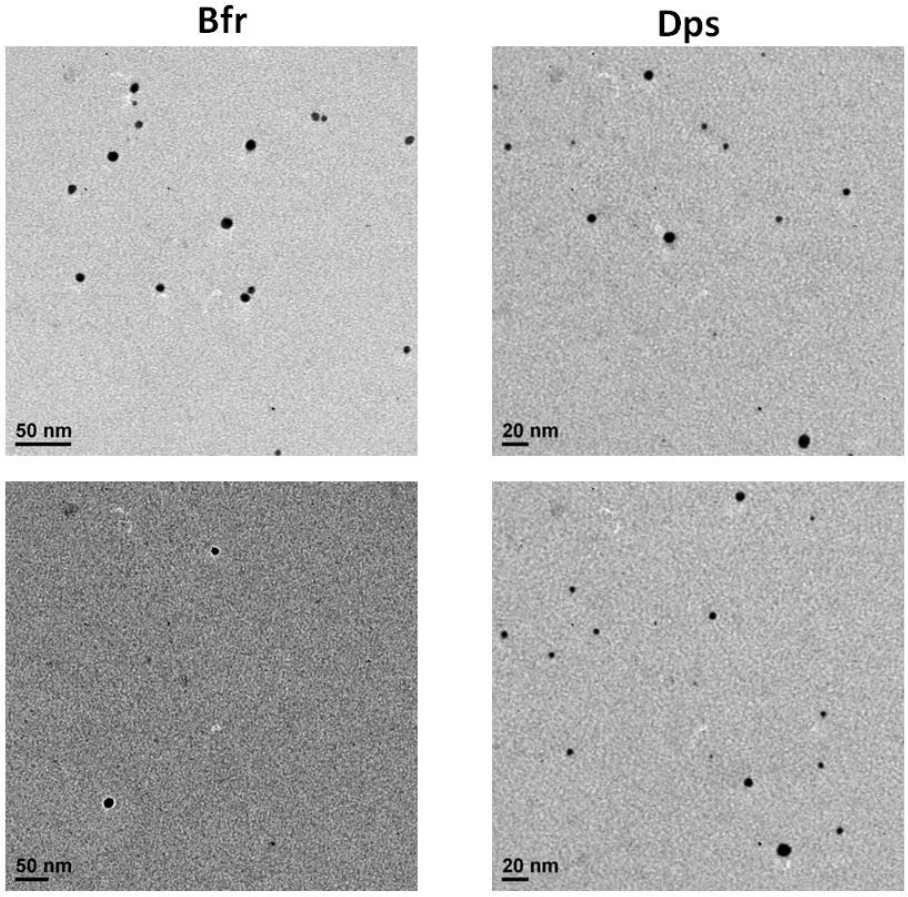


Figure 2.7 – Transmission electron microscopy of semi optimised nanoparticle genesis condition from the SEC elution volume corresponding to the Bfr and Dps nanocages seen in figure 2.4. Nanoparticle experiment from trial 16, used 1 ml of 1mg/ml protein in Tris buffer (50 mM Tris.HCl, 50 mM NaCl, pH 7) incubated with 20 µl of 0.1 M $HAuCl_4$ buffered in Tris (50 mM Tris.HCl, pH7.8) for 3 hrs, at room temp. 20 µl of 0.1 M $NaCNBH_3$ buffered in Tris (50 mM Tris.HCl, pH7.8) was then added and was incubated for 3 hrs, at room temp followed by desalting. An additional 10 µl of 0.1 M $HAuCl_4$ buffered in Tris (50 mM Tris.HCl, pH7.8) and 30 µl of 0.1 M ascorbic acid was added and incubated, overnight at room temp. After concentration, 0.5 ml of the remaining sample purified via SEC and this sample was subject to TEM using 1% uranyl acetate as a negative stain.

Buffering the reagents and the use of a weaker reducing agent did seem to have an effect on the species observed under SEC. With the change in pH seemingly being resolved through reagent buffering, further trails analysed the concentration and volume of the gold source and the first reducing agent. The concentration of the $HAuCl_4$ was dropped to 0.01 M but, to ensure enough gold was present in the final solution and was added in a way to prevent

protein aggregation, small volumes were added multiple times (5 x 20 µl additions ten minutes apart) while keeping the first reducing agent the same as above (20 µl addition of 0.1 M $NaCNBH_3$) (Trial 17). Smaller volumes of $HAuCl_4$ solution were trialled including 100 µl of 0.05 M (Trial 18) and 20 µ of 0.1 M (Trial 19) while keeping the first reducing agent the same as above (20 µl addition of 0.1 M $NaCNBH_3$). This condition was repeated but with an increased volume of first reducing agent (100 µl of 0.1 M $NaCNBH_3$) with both a 5 x 20 µl of 0.05 M $HAuCl_4$ solution addition (Trial 20) and a 100 µl of 0.05 M $HAuCl_4$ solution addition (Trial 21). However, under all conditions, SEC analysis showed a continual signal across a wide elution volume similar to figure 2.6.

The only reagent which has so far not been examined was the second reducing agent, ascorbic acid. As the name suggests, this too is acidic and could be playing a role in the aggregation of the proteins. Initially this reagent was overlook as it is found in cells and therefore should be biocompatible. All further trials were conducted with all reagents buffered in Tris (50 mM Tris.HCl, pH 7.8). Experiments were performed where the concentration and volume of the first (100 µl of 0.1 M $NaCNBH_3$ in 50 mM Tris.HCl, pH 7.8 buffer) and second (50 µl of 0.1 M ascorbic acid in 50 mM Tris.HCl, pH 7.8 buffer) reducing agents were kept constant, but the concentration and volume of the $HAuCl_4$ solution from both additions was varied. For the first $HAuCl_4$ addition 5 x 20 µl (Trial 22) 5 x 40 µl (Trial 23) and 1 x 100 µl (Trial 24) of gold at 0.05 M was added with 10 µl of 0.1 M gold added in the second addition for all three trials. Further optimisation repeated the last three trials but using 100 µl of 0.05 M $HAuCl_4$ for the second additions while varying the first gold addition to 5 x 20 µl (Trial 25) 5 x 40 µl (Trial 26) and 1 x 100 µl (Trial 27) of at 0.05 M $HAuCl_4$.

SEC analysis showed the presence of not just a UV 280 nm signal across the nanocage elution volume but a well-defined peak corresponding relatively well with the nanocages of both Bfr and Dps for all trials 22 - 27 with the SEC from trial 27 represented below (Figure 2.8).

Well defined peaks were seen for both proteins, at the expected nanocage elution volume, with corresponding SPR signals. This co-elution suggested the presence of intact protein nanocages either attached to size selective nanoparticles or with nanoparticles encapsulated within them. The presence of only a small aggregation peak from Bfr and Dps suggests that this new mild condition is optimal for nanocage survival.

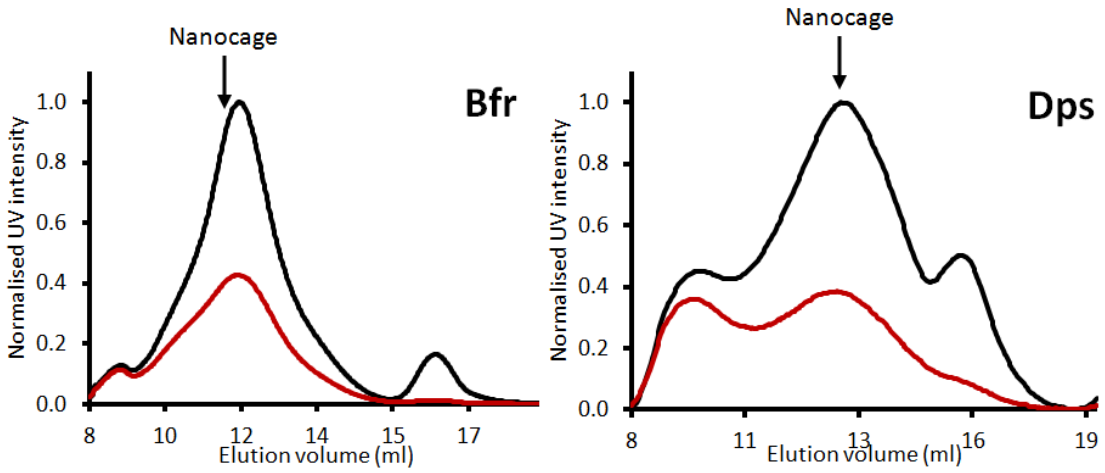


Figure 2.8 – Further optimised gold nanoparticle genesis conditions buffering all reagents in Tris buffer (50 mM Tris, pH 7.8.). Size exclusion chromatograms of purified Bfr (left) and Dps (right) after being subjected to nanoparticle generation conditions, protein 280 nm (black) and SPR 530 nm (red). Nanoparticle experiment from trial 27, used 1 ml of 1mg/ml protein in Tris buffer (50 mM Tris.HCl, 50 mM NaCl, pH 7) incubated with 100 µl of 0.05 M $HAuCl_4$ buffered in Tris (50 mM Tris.HCl, pH7.8) for 3 hrs, at room temp. 100 µl of 0.1 M $NaCNBH_3$ buffered in Tris (50 mM Tris.HCl, pH7.8) was then added and was incubated for 3 hrs, at room temp followed by desalting. An additional 100 µl of 0.05 M $HAuCl_4$ buffered in Tris (50 mM Tris.HCl, pH7.8) and 50 µl of 0.1 M ascorbic acid buffered in Tris (50 mM Tris.HCl, pH7.8) was added and incubated, overnight at room temp. After concentration, 0.5 ml of the remaining sample was injected. Each sample was run once.

To assess whether or not intact nanocages are present, a sample from the elution volumes corresponding for nanocages were collected and examined via TEM (Figure 2.9). The presence of empty nanocages was very evident for Bfr showing that this method does not completely lead to protein aggregation. Nanoparticles were also observed, but with two distinct sizes. The larger sized nanoparticles that show the presence of a white protein ring around them are the size expected for the internal cavity of Bfr. However, the smaller size which appear more frequently, tend to also have pale rings, or protein shells around them and could be nanoparticles that have not yet grown to fill the cavity. This could be because of the

change to a weaker reducing agent. However the presence of intact nanocages, suggests that the pale rings seen around the nanoparticles are nanocages and not protein aggregates. Although this is promising, there are clearly nanoparticles that are present outside of the nanocages. The Dps TEM images present a different picture. No intact nanocages can be observed even although a clear peak in the SEC (Figure 2.8) at the correct elution volume for Dps can be seen. Nanoparticles of two sizes are again observed. A larger nanoparticle, which is too large to be from the internal cavity of Dps and a smaller particle whose size is similar to that of the cavity, were seen. Further TEM analysis with a higher intensity electron beam could be useful to ascertain the presence of intact Dps nanocages.

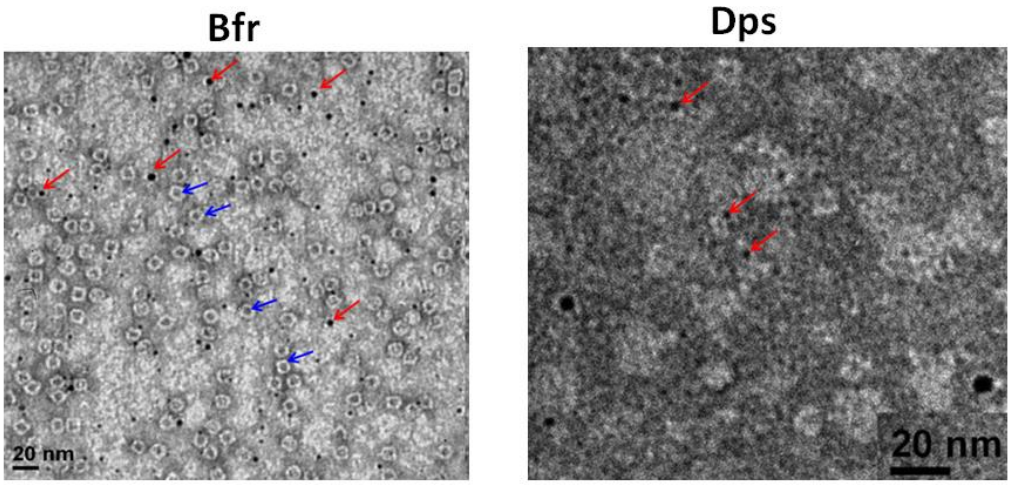


Figure 2.9 – Transmission electron microscopy of further optimised nanoparticle genesis condition from the SEC elution volume corresponding to the Bfr and Dps nanocages seen in figure 2.8. (left) Bfr, highlighting both empty but intact nanocages (blue arrow), and nanoparticles with pale rings of a size similar to the Bfr inner cavity (red arrow). (right) Dps, highlighting nanoparticles with pale rings of the size constant with the Dps inner cavity (red arrow). Nanoparticle experiment from trial 27, used 1 ml of 1mg/ml protein in Tris buffer (50 mM Tris.HCl, 50 mM NaCl, pH 7) incubated with 100 µl of 0.05 M $HAuCl_4$ buffered in Tris (50 mM Tris.HCl, pH7.8) for 3 hrs, at room temp. 100 µl of 0.1 M $NaCNBH_3$ buffered in Tris (50 mM Tris.HCl, pH7.8) was then added and was incubated for 3 hrs, at room temp followed by desalting. An additional 100 µl of 0.05 M $HAuCl_4$ buffered in Tris (50 mM Tris.HCl, pH7.8) and 50 µl of 0.1 M ascorbic acid buffered in Tris (50 mM Tris.HCl, pH7.8) was added and incubated, overnight at room temp. After concentration, 0.5 ml of the remaining sample purified via SEC and this sample was subject to TEM using 1% uranyl acetate as a negative stain.

While SEC analysis of this semi-optimised condition from trial 27, showed the presence of nanocages, which for Bfr was verified through TEM, work is still needed to explore further conditions that improve on this initial data for Bfr and for Dps.

Due to time constraints, further optimisation was only conceived. By the presence of small nanoparticles that did not fill the cavity, it is apparent that the growth of the nanoparticle is insufficient. Therefore further conditions could be analysed that enhance this stage. Incubation times, the volume and concentrations of both the first and second reducing agents could be increased as well as the amount of gold added during the second reduction step. A range of other reducing agents could also be examined for their potential to reduce gold in a manner that does not aggregate the proteins. However, the examination of so many alterable conditions using this current system is time and resource expensive. If a quicker method can be obtained that can screen for protein stability in the presence of different reagents at varying concentrations, nanoparticle formation could be optimised faster. This screen is further discussed in Chapters 4 and 5 and used to examine nanoparticle formation conditions in Chapter 5.

2.2.2. Radioactive labelling of a apo-ferritin for *in vivo* localisation in mice

Protein nanocages are potential candidates for a range of applications (see Chapter 1). One such usage would be in targeted drug delivery. Through modification of the protein sequence or through fusions, protein nanocages can be made to target areas of the body, thus, in combination with its cargo, be used as a targeted delivery agents. In order to fully appreciate the ability of ferritin nanocages for such applications, some fundamental questions first need to be asked. Among these are “How to determine the targeting?” and “Where do ferritin have a background targeting preference?” We set out to answer these questions in this chapter. We designed a strategy to radiolabel the readily available and highly stable HsFn L-chain. The resulting radioactive complexes were then injected into mice and their location

monitored with single-photon emission computed tomography and X-ray computed tomography (SPECT/CT) over the course of several days. The data obtained from this investigation would form a foundation for targeted drug delivery using ferritins.

2.2.2.1 Conjugation of CHX-A'-DTPA ligand to horse spleen ferritin and radiolabelling with ^{111}In

Light chain horse spleen ferritin is a highly stable, readily available and non-catalytically active protein nanocage. It was chosen as a robust model system for this investigation. The ligand *N*-[(R)-2-Amino-3-(*p*-isothiocyanato-phenyl) propyl]-*trans*-(*S,S*)-cyclohexane-1,2-diamine-*N,N,N',N",N"*-pentaacetic acid (CHX-A'-DTPA, Macrocyclics) (Figure 2.10) has been used often by our collaborator's laboratory, and was chosen for this investigation due to their expertise and its high binding affinity to ^{111}Indium. Isothiocyanate groups are able to react with the amine group of lysine side chains. HsFn L-chain has nine lysine residues (K59, K68, K84, K98, K105, K140, K143, K144 and K173). While several are either buried (K59, K140, K143 and K144) or inside the potentially ligand inaccessible cavity (K68 and K173), some are surface exposed and accessible (K84, K98 and K105) (PDB:2W0O, see Appendix 2.1 for full sequence). These residues are expected to be the targets for conjugation to the CHX-A'-DTPA ligand.

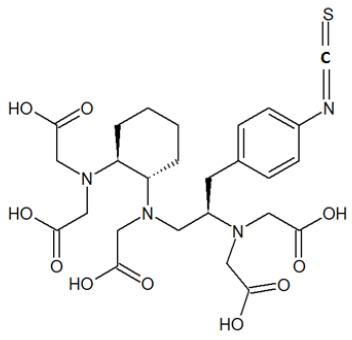


Figure 2.10 – Chemical structure of *N*-[(R)-2-Amino-3-(*p*-isothiocyanato-phenyl) propyl]-*trans*-(*S,S*)-cyclohexane-1,2-diamine-*N,N,N',N",N"*-pentaacetic acid (CHX-A'-DTPA).

CHX-A'-DTPA was conjugated to horse spleen ferritin through a simple overnight incubation and then characterised by the ability of the resulting complex to bind to radioactive ^{111}In. HPLC with dual protein (280 nm) and radiation monitoring can be used to generate co-elution profiles to identify protein species associated with radioactivity. HPLC, using a size exclusion chromatography column, of HsFn alone was run as a standard (Figure 2.11). The 280 nm channel shows a single peak at 8 ml and with no associated radiation.

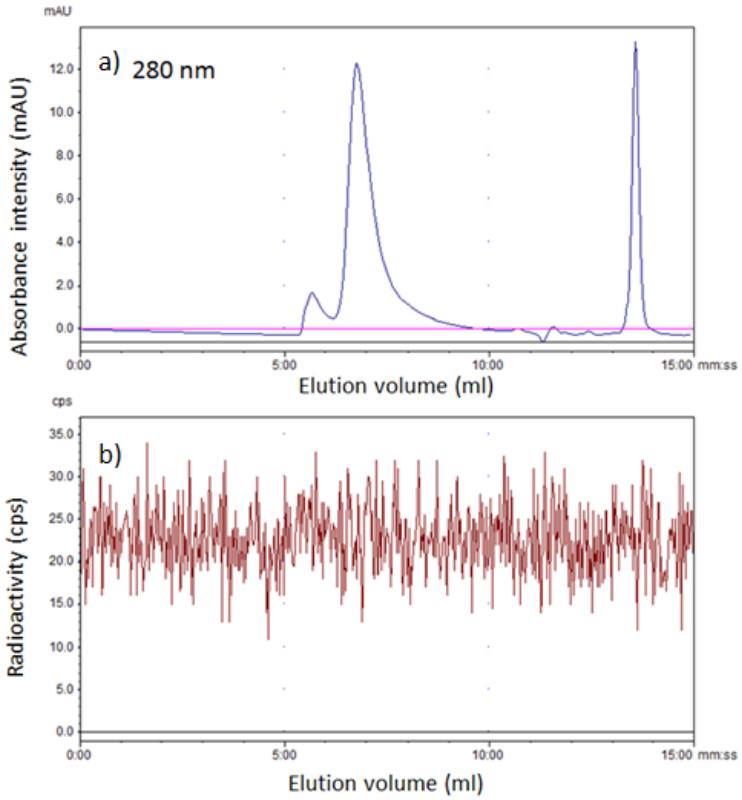


Figure 2.11 – Analysis of the radioactivity of the unconjugated HsFn protein nanocage. Size exclusion chromatography column used in conjunction with HPLC. Protein monitored at 280 nm (10 µl of 5 mg/ml protein sample in 0.1 M ammonium acetate) run at 1 ml/min (running buffer – 0.1 M Na_2HPO_4, 0.001 M EDTA, pH 7.5). Sample was run once. a) HPLC of sample through size exclusion column, b) radioactivty.

As ferritins are known to bind metal ions, it was important to determine whether HsFn can bind ^{111}In. If binding or trapping did occur, then this could show false positives for the successful incorporation of CHX-A'-DTPA, but, ferritin would most likely have a much lower affinity for ^{111}In than CHX-A'-DTPA allowing for its undesirable release during *in vivo* analysis. Therefore, unconjugated HsFn was incubated with ^{111}In and analysed by HPLC (Figure 2.12).

The protein channel showed a similar profile to the standard (Figure 2.9). It should be noted that the buffer used in this analysis contains ethylenediaminetetraacetic acid (EDTA) which sequesters ^{111}In. The resulting complex elutes from this column at 13 ml. The radioactive channel of the protein interjection only showed this peak. Importantly, no radiation co-eluted with the ferritin suggesting that it does not bind ^{111}In when it is not conjugated to CHX-A'-DTPA.

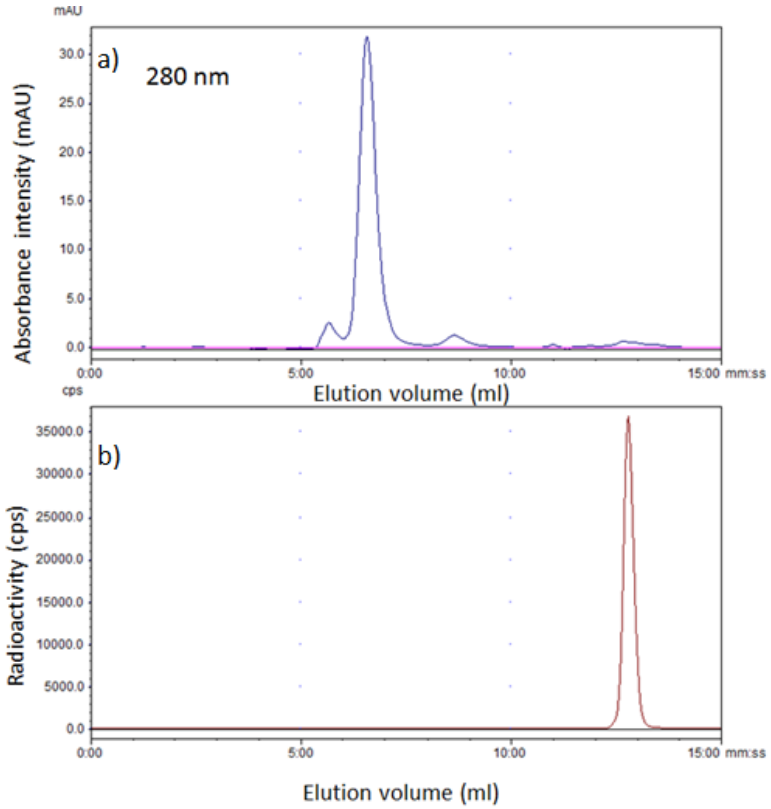


Figure 2.12 – Analysis of the radioactivity from a unconjugated HsFn protein nanocage after 30 min incubation with ^{111}In. Size exclusion chromatography column used in conjunction with HPLC. Protein monitored at 280 nm (10 µl of 5 mg/ml in 0.1 M ammonium acetate, pH 6, mixed with 5 µl of ^{111}In with 128 Mbq of activity) run at 1 ml/min (running buffer – 0.1 M Na_2HPO_4, 0.001 M EDTA, pH 7.5). Sample was run once. a) HPLC of sample through size exclusion column, b) radioactivty.

With the controls established, the CHX-A'-DTPA-ferritin complex was subjected to the same HPLC analysis after incubation with ^{111}In (Figure 2.13). The HsFn nanocage peak was evident in the protein channel, indicating that the majority of the protein survives the conjugation. Importantly, this peak co-elutes with the radioactivity, demonstrating that the conjugation reaction was successful and the ligand can be successfully loaded with ^{111}In. With

the complex in hand we were prepared to use it to monitor the position of exogenous ferritins inside of a complex organism through *in vivo* radioactive monitoring techniques (SPECT/CT).

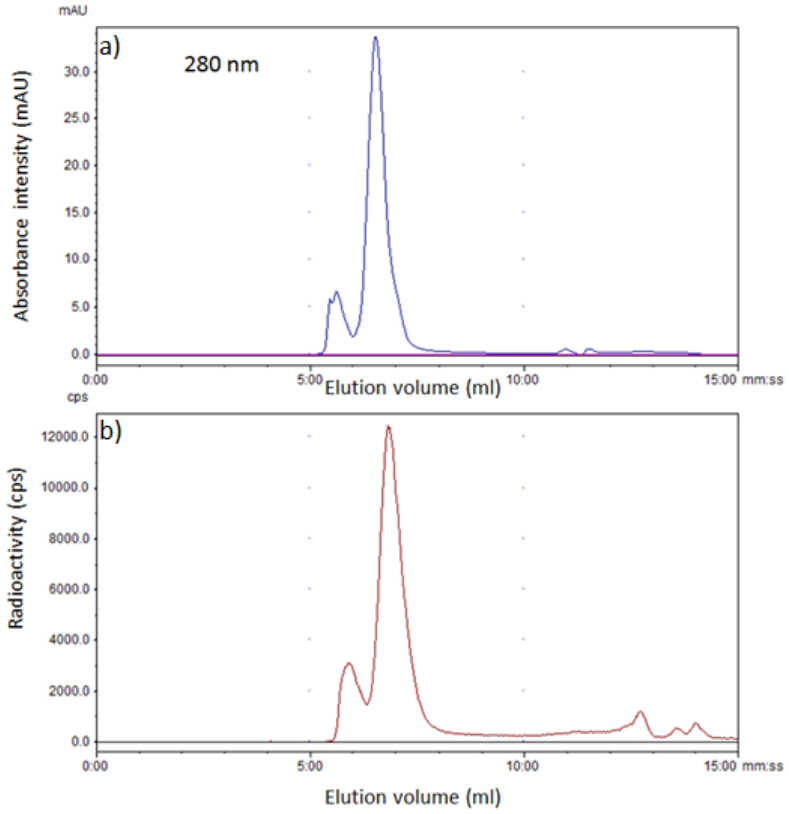


Figure 2.13 – Analysis of the radioactivity of the protein nanocage HsFn after being conjugated to CHX-A'-DTPA and incubated for 30 min with ^{111}In. Size exclusion chromatography column used in conjunction with HPLC. Protein monitored at 280 nm (10 µl of 5 mg/ml in 0.1 M ammonium acetate, pH 6, mixed with 5 µl of ^{111}In with 128 Mbq of activity) run at 1 ml/min (running buffer – 0.1 M Na_2HPO_4, 0.001 M EDTA, pH 7.5). Sample was run once. a) HPLC of sample through size exclusion column, b) radioactivty.

2.2.2.3. *In vivo* localisation of exogenous ferritin in mice

The CHX-A'-DTPA-ferritin complex was radioactively labelled with ^{111}In, injected in the tail vein of three mice and a SPECT/CT scan was performed at 0, 3, 24 and 48 hours post injection (Figure 2.14 and Figure 2.15). For the 0 and 3 hour post injection images, a large radioactive response was detected in the liver, intestine and bladder, corresponding to normal routes of expulsion from the body. After 24 hours, the liver was still highly active but the majority of the mice had much lower responses. Interestingly two additional pairs of 'hot spots' were seen with much higher activity than the surrounding tissue (Figure 2.12 areas 1 and 2) and even after 48 hours these were still evident. While the filtration and excretory

systems would be expected to be active, the hot spots were intriguing as they might suggest some specific targeting by ferritin even without any appended targeting functionality

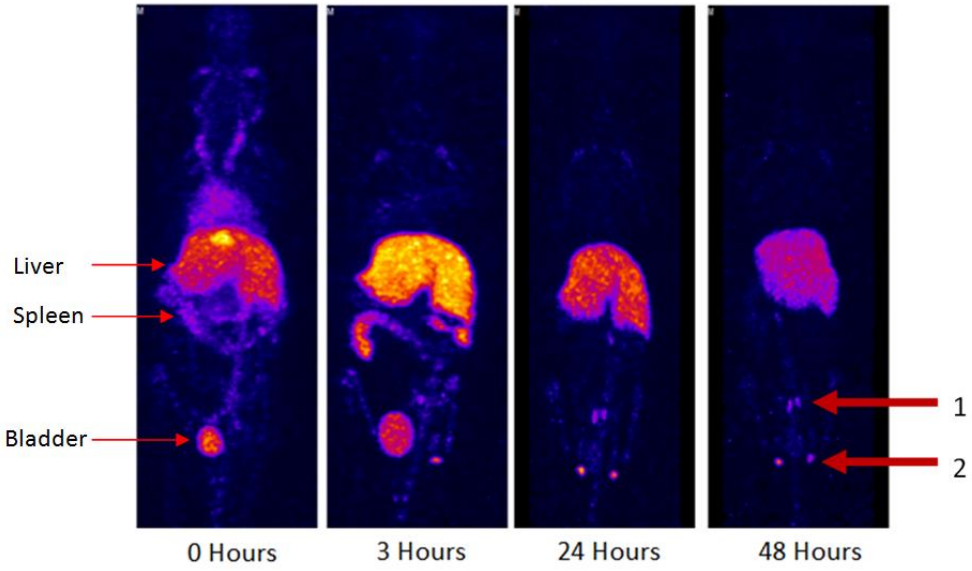


Figure 2.14 – Two-dimensional SPECT image for a mouse injected with ^{111}In CHX-A'-DTPA-ferritin complex after 0, 3, 24 and 48 hours. Each injection contained 0.1 mg of protein incubated with 25 µl of ^{111}In with 193 Mbq of activity for 30 min, then made up to 100 µl with saline. Each syringe contained 16 Mbq of activity after preparation, with 0.1047 g (105.4 µl), 0.1024 g (102.4 µl), and 0.1078 g (103.8 µl) being injected into the tail vein of mouse 1, 2 and 3.

The combined SPECT reconstruction and the CT scan can be used to direct dissection as a first step toward identifying these hotspots (Figure 2.14). The mice were culled and their organs harvested and a mass that corresponded to the shape and location of area 1 from figure 2.12 was collected and its identity was suggested to be that of dorsal lymph nodes. The second area, labelled as 2 in figure 2.14, was not identified and no organ in that location corresponding to that shape could be harvested from the mice. Its location as seen from 3D reconstructions suggests it could be located between the joints in the hip bone, but this could not be confirmed and further analysis is required (Figure 2.15).

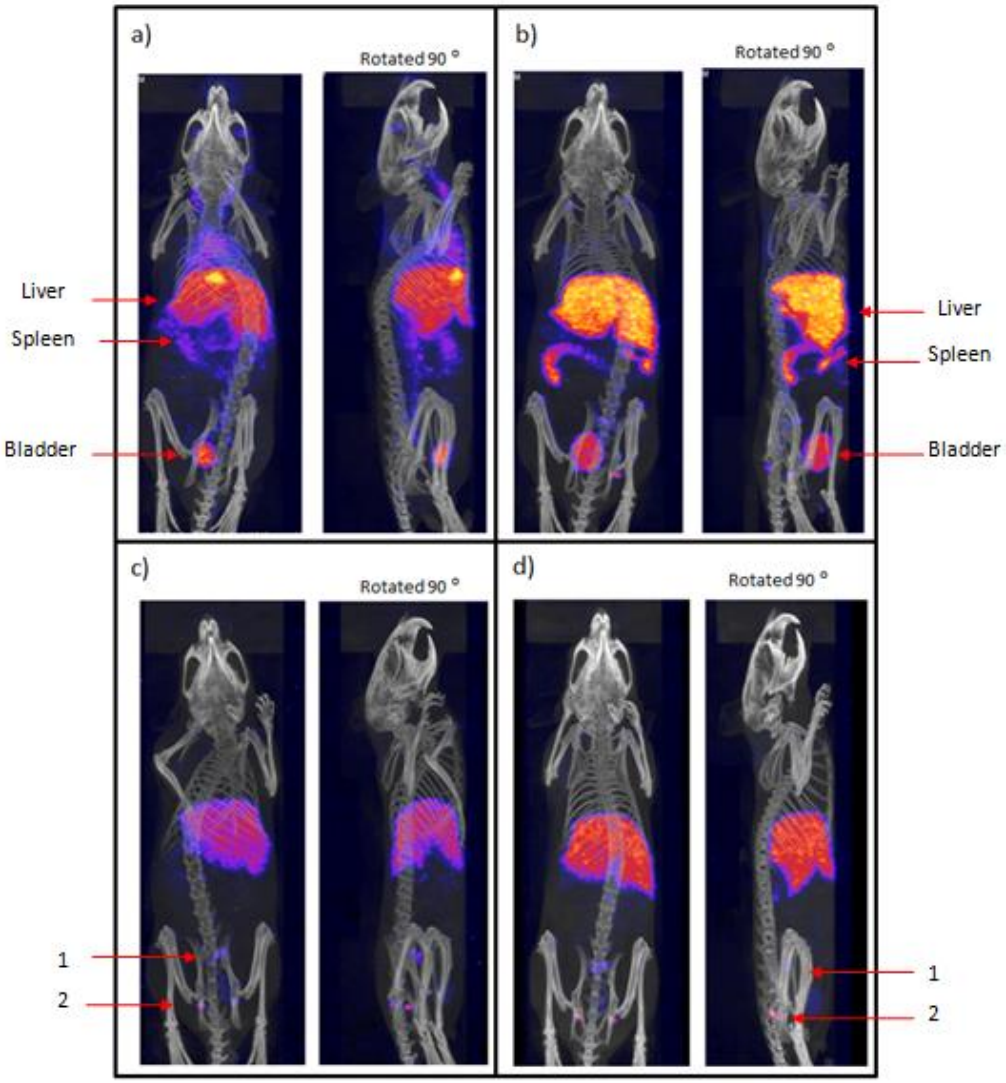


Figure 2.15 – SPECT/CT reconstruction of a mouse injected with ^{111}In CHX-A'-DTPA-ferritin complex after (a) 0 hours, (b) 3 hours, (c) 24 hours and (d) 48 hours. Each injection contained 0.1 mg of protein incubated with 25 µl of ^{111}In with 193 Mbq of activity for 30 min, then made up to 100 µl with saline. Each syringe contained 16 Mbq of activity after preparation, with 0.1047 g (105.4 µl), 0.1024 g (102.4 µl), and 0.1078 g (103.8 µl) being injected into the tail vein of mouse 1, 2 and 3. Bladder was not removed during image reconstruction.

The dissected organs were analysed and their activities were determined with respect to their weights (Figure 2.16). As expected from the SPECT reconstructions, the liver showed high activity as did the spleen. Area 1 from Figure 2.14, shows very high activity for its size and has clearly been labelled strongly by the CHX-A'-DTPA-ferritin complex. There also is slight activity in bone.

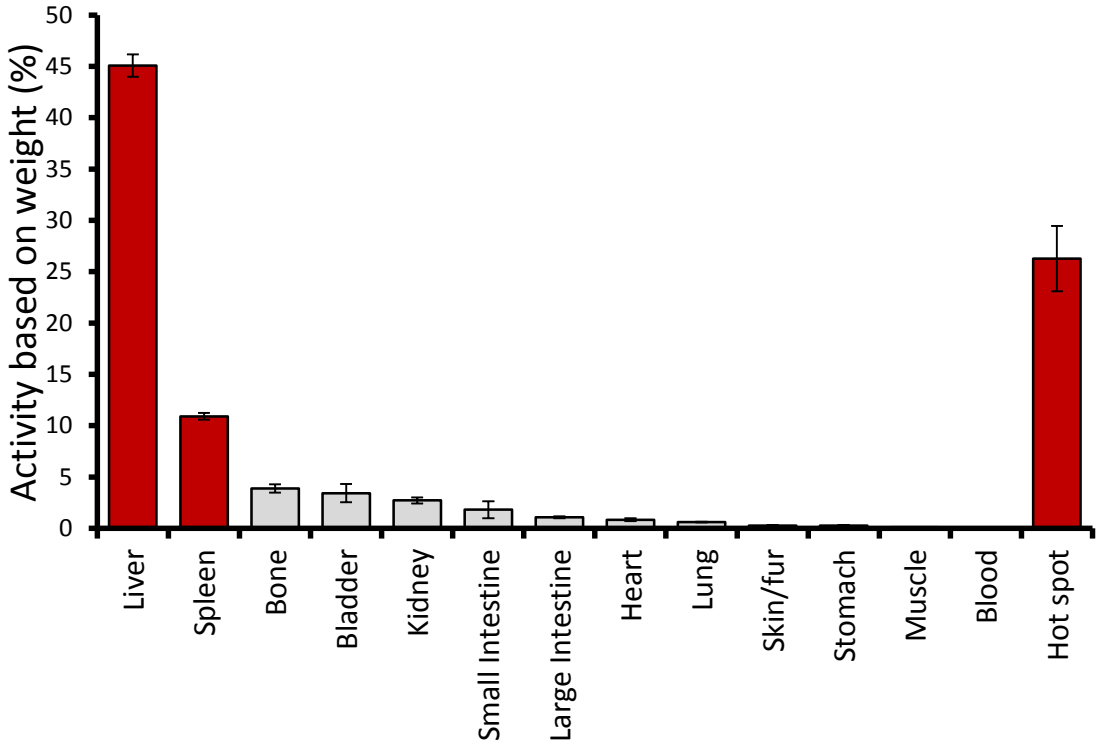


Figure 2.16 – Comparison of radioactivity based on weight to the control (see Appendix 2.8) of organs harvested from the mice after 48 hours post injection of the ^{111}In labelled CHX-A'-DTPA-ferritin complex. 'Hot spot' is area 1 from Figure 2.12. Data is averaged from the organs of three mice.

Further controls are required to ensure that this native targetting is not due to just the ligand, but is infact a product of the ferritin. However, if that data presented here is found to be due to the ferritin and not the ligand, then this inital data could suggest that even without targeting functionality added, HsFn could target a number of locations within a mouse. Labelling of the spleen might suggest a natural uptake mechanism to this iron rich organ for iron binding proteins. The low but significant binding to bone and the highly selective labelling to two 'hot spot' areas, warrants further investigation. Even if these areas are not bio-medically interesting targets, this data can be used as a first step to characterizing a drug delivery and imaging method using ferritins with fused to specific targeting moieties.

2.3. Conclusions

Our laboratories previously published method for the formation of mono-dispersed gold nanoparticles inside of light chain horse spleen ferritin was expanded to two other ferritins with the aim of showing the versatility and breath of this technique. These ferritins, differing in size and symmetry, and both from bacteria behaved differently than HsFn in the nanoparticle generating conditions with both proteins completely aggregating. The conditions were explored more deeply and the problem was determined to be a combination of pH and the strong reducing agent used in the first reduction step. Through a combination of using a weaker reducing agent, and buffering all solutions, the survival of both proteins was enhanced. However, under these conditions optimised for protein stability, the formation of nanoparticles was less than optimal, with different particle sizes observed for both Bfr and Dps. This may be a consequence of lowering the strength of the first reducing agent. While this work is a step forward in generalizing our method toward other protein nanocages, especially those less robust than HsFn, this technique must still be further optimised to reliably produce protein nanocage encapsulated gold nanoparticles.

This investigation did highlight some challenges to working with nanocages. Applying them to various nanotechnologies can sometimes be hindered if their fundamental physical and chemical properties are not fully understood or characterised. A step in this direction is addressed in Chapter 3. Furthermore, although several new conditions were assayed for their potential to not only form nanoparticles but also to maintain the structural integrity of the proteins, this process was slow and relied upon traditional low throughput techniques such as SEC. If a method could be found that allows for the rapid assessment of protein quaternary structure without relying on slow characterisation techniques, the throughput could be expanded, and many more conditions and reagents could be screened in a much more rapid manner. Not only could such a method help to swiftly expand the nanoparticle generating

methodology, it could also be used as a tool to better understand the fundamentals of quaternary structure of nanocages, which ultimately would aid in their utilisation. The development of such a method is explored in Chapters 4 and 5, and its utilisation for engineering more robust protein nanocages is discussed in Chapter 6. An important take-home lesson from our work with the nanoparticles is that HsFn is clearly much more robust than Bfr and Dps. As the long term aim of the project described in this chapter is to explore ways to exploit ferritin nanocages for novel applications, it became apparent, that using a highly stable ferritin would provide the most optimal route. Thus, for the second half of this chapter, where we applied ferritins to imaging, we employed HsFn.

Another envisioned application of protein nanocages is as delivery and transport vessels. It is imagined that the cavities could be loaded with cargo, be it small molecule drugs, genomic material or even nanoparticles, and the "package" would be targeted to certain tissues within the body. However, before targeting studies can take place, initial research into the suitability of ferritins for these applications must be determined. Therefore we sought to establish the background targeting of ferritins in living mice. We conjugated HsFn to a radioactive reporter to allow visualisation of the ferritin in living mice using SPECT/CT. The conjugation was successful most likely due to the robust nature of this particular ferritin and readily visualized particular organs in the mouse. We did see accumulation in the liver, spleen and two "hot spots" one of which may be the dorsal lymph nodes. This alone could suggest its use as a targeting agent, or this data could be used as background for applications with more specific targeting. Taken together, these data indicate that HsFn could be a suitable tool for the further development of targeted delivery applications.

2.4. Methods

2.4.1. Generation of ferritin encapsulated nanoparticles

2.4.1.1. Protein production and purification

Vectors (pET-32b) containing the genes coding for wild type Bfr and Dps (see Sequence Appendix 2 for full sequences) were electroporated into BL21 *E.coli* cells and plated on LB plates (50 µl/ml of carbenicillin). Selected colonies were then grown in LB (5 ml, 37 °C, overnight) as pre-culture which was later added to LB (500 ml) and grown (37 °C) until an $O.D_{600}$ of 0.6. Protein expression was then induced by the addition of IPTG (final concentration of 400 mM) and the cultures were further incubated (3 h, 30 °C). The cells were isolated by centrifugation (10,000 rpm, 20 min at 4 °C). The cell pellet was resuspended with lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 40 mM Imidazole, pH 8) and sonicated (15 min, pulsed 10 s off then on, with a Sonics vibra cell sonicator). The protein solutions were clarified by centrifugation (10,000 rpm, 20 min at 4 °C) and then filtered (Pall, 0.2 µm).

The protein was purified via affinity chromatography via the His_6 tag with a GE 5 ml Histrap FF Column using wash buffer (30 mM NaH_2PO_4, 300 mM NaCl, 40 mM Imidazole, pH 7.6). The proteins of interest were eluted from the Histrap using elution buffer (30 mM $NaHPO_4$, 300 mM NaCl, 500 mM Imidazole, pH 7.6) and then digested. Enterokinase digestion (NEB 2 µg/ml) was performed to cleave off the peptide tag from the protein of interest followed by a second Histrap (GE, Histrap FF, 5 ml, (wash buffer-40 mM Imidazole, 50 mM NaH_2PO_4, 300 mM NaCl, pH 7.4), (elution buffer-500 mM Imidazole, 50 mM NaH_2PO_4, 300 mM NaCl, pH 7.4)) to remove the tag from solution. The protein solutions were further purified by size exclusion chromatography (GE Hiload 16/60 Superdex with running buffer (50 mM Tris.HCl, 50 mM NaCl, pH 7.8.)). The degree of purification was assessed with SDS PAGE (see Appendix 2.2).

2.4.1.2. Iron removal from pure protein

The purified proteins underwent dialysis into 'apo buffer' (1% thioglycolic acid, 1% 2'2' bipyridine, pH 4.5) for 24 hours to remove any iron[39]. This buffer was then changed to a Tris buffer (50 mM Tris.HCl, 50 mM NaCl, pH 7.8) and left for a further 24 hours. After dialysis the protein sample was adjusted to 1 mg/ml ready for nanoparticle experiments.

2.4.1.3. Initial conditions for the formation of nanoparticles inside unmodified apo-ferritins[1]

For control experiments, purchased apo-horse spleen ferritin light chain (HsFn, Sigma, 5 mg/ml stock) was purified via SEC (1 ml, running buffer 50 mM Tris, 50 NaCl, pH 7.8, GE Superdex 200 10/300 GL) then diluted for nanoparticle experiments (1 mg/ml).

The initial method for the creation of nanoparticles as previously described required the addition of a first gold source (22.6 µl of a 0.1 M $HAuCl_4$ solution in water) to 1 ml of a 1 mg/ml protein sample. This solution was incubated (3 h, room temp). The sample was then desalted using a GE 5 ml HiTrap desalting column and Tris buffer (50 mM Tris, 50 mM NaCl, pH 7.8) to remove any un-encapsulated gold. The protein was collected (1 ml) and the first reducing agent was added (20 µl of 0.1 M $NaBH_4$ solution in water) and the resulting solution was incubated (3 h, room temp) while shaking. After initial reduction, a second gold source was added (10 µl of a 0.1 M $HAuCl_4$ solution in water) followed by a second reducing agent (30 µl of a 0.1 M Ascorbic acid solution in water) followed by incubation (overnight, room temp). Samples were centrifuged (10 min, 10,000 rpm) before characterisation.

All alterations to the above method including changes to the concentration, volume and buffer for the first and second gold sources and reducing agents, are described in Appendix 2.6.

2.4.1.4. Screening of buffer conditions for the reduction of gold without protein

To analyse the best buffering conditions for the reduction of gold a medium throughput screen was envisioned. To a stock buffer (1 ml of either dH_2O unbuffered; or 50 mM Tris.HCl, 50 mM NaCl, pH 7.8; or 50 mM Na_2HPO_4, 50 mM NaCl, pH 7) gold solution was added (20 µl in either dH_2O unbuffered; or 50 mM Tris.HCl, 50 mM NaCl, pH 7.8; or 50 mM Na_2HPO_4, 50 mM NaCl, pH 7), shaken, followed by the introduction of a reducing agent (20 µl in either dH_2O unbuffered; or 50 mM Tris.HCl, 50 mM NaCl, pH 7.8; or 50 mM Na_2HPO_4, 50 mM NaCl, pH 7). This was incubated (overnight, room temperature) and its colour analysed.

2.4.1.5. Modified method for the formation of nanoparticles inside unmodified apo-ferritins

This is the most successful method to date for the formation of nanoparticles inside both Dps and Bfr, while causing the least damage to the protein nanocages. To 1 mg of 1 mg/ml protein sample the first gold source was added (100 µl of 0.05 mM $HAuCl_4$ in Tris buffer (50 mM Tris.HCl, pH 7.8)) and the solution was incubated (2 h, rom temp). This sample was desalted using a GE 5 ml HiTrap desalting column and Tris buffer (50 mM Tris, 50 mM NaCl, pH 7.8) to remove any un-encapsulated gold. The protein peak was collected (1 ml) and the first reducing agent was added (100 µl of 0.1 M $NaCNBH_3$ in Tris buffer (50 mM Tris.HCl, pH 7.8)) and the solution was incubated (3 h, room temp) while shaking. After incubation, a second gold source was added (100 µl of 0.05 mM $HAuCl_4$ in Tris buffer (50 mM Tris.HCl, pH 7.8)), as well as a second reducing agent (50 µl of 0.1 mM ascorbic acid in Tris buffer (50 mM Tris.HCl, pH 7.8)) and this was incubated (overnight, at room temperature) before characterisation.

2.4.1.6. Surveying of nanoparticle formation via Size exclusion chromatography (SEC)

All samples were analysed for both protein stability (280 nm) and nanoparticle formation (520 nm) via SEC. The samples (0.5 mg/ml) in Tris buffer (50 mM Tris.HCl, 50 mM NaCl, pH 7.8) was injected (0.5 ml) on to the column (GE Superdex 200 10/300 GL) at 0.5

ml/min. The column was calibrated using six proteins as standards (GE Biosystems Calibration Kit).

2.4.1.7. Surveying of nanoparticle formation via Transmission electron microscopy (TEM)

TEM was performed on Jeol JEM-1400, operating at 100keV, electron microscope set at 200 KeV. Proteins were immobilized on Formvar/carbon coated 3.05 mm copper grids (TAAB) and negatively stained with 1% Uranyl acetate using 0.05 mg/ml protein concentrations.

2.4.1.8. Circular Dichroism Spectroscopy (CD)

All proteins were purified into phosphate buffer (50 mM Na_2HPO_4, 50 mM NaCl, pH 7). This experiment was performed on JASCO J-810 spectropolarimeter from 250 nm to 200 nm with a protein concentration of 0.2 mg/ml with a path length of 0.5 mm. Thermal melts were performed on all purified proteins (0.2 mg/ml in phosphate buffer, 50 mM Na_2HPO_4, 50 mM NaCl, pH 7) in a range of 4 to 100 °C[40, 41] (see Appendix 2.3).

2.4.2. Imaging the localization of the ^{111}In-CHX-A'-DTPA-Ferritin complex

2.4.2.1. Conjugation of CHX-A'-DTPA to Horse spleen ferritin (HsFn)

Purchased apo-horse spleen ferritin light chain (HsFn, Sigma, 5 mg/ml stock) was purified via SEC (1 ml, running buffer 50 mM Tris, 50 NaCl, pH 7.8, GE Superdex 200 10/300 GL). This protein sample was concentrated (using a 100 kDa cut off ultrafiltration tube (Millipore)) followed by the addition of EDTA (50 µl of 0.01 M EDTA solution, incubated for 30 min) to ensure that this sample is free from any metal contaminates, and then buffer exchanged (0.1 M HEPES buffer pH 8.9) via several washing steps to remove all EDTA. This solution was diluted (1 ml of 4 mg/ml protein) to which the ligand CHX-A'-DTPA (3.4 mg, Macrocyclics) was added followed by agitated incubation (overnight at room temperature). This sample underwent buffer exchange via several washing steps (using a 100 kDa cut off

ultrafiltration tube (Millipore) into 0.1 M ammonium acetate, pH 6, free from metal contaminates). The number of washing steps depended on the noise seen in HPLC, for this experiment the sample was washed 12 times. The final sample had a concentration of 3.4 mg/ml and a volume of 1 ml which was a 68% recovery.

2.4.2.2. Radioactive labelling of CHX-A'-DTPA-Ferritin complex and HPLC characterisation

HPLC with an inline radio nucleotide detector was used to verify the state of the protein after the reaction and whether or not the bio-conjugation worked and the sample can now bind to the radioactive isotope.

To the ferritin sample (10 µl of 3.4 mg/ml in 0.1 M ammonium acetate, pH 6) the radioactive isotope was added (5 µl of ^{111}In with 128 Mbq of activity, Coridien, Petten) and the resulting solution was incubated (30 min at room temperature). This sample was injected into a Bio-sep-SEC-S2000 HPLC column (10 µl of sample, 1 ml/min, running buffer – 0.1 M Na_2HPO_4, 0.001 M EDTA, pH 7.5) and analysed for the presence of protein (280 nm) and radioactivity.

2.4.2.3. *In vivo* SPECT imaging

For *in vivo* experiments, three mice were used (Balb/C, female, 8-12 weeks old). In total, four injections were prepared, three for the mice, one as a control (each sample contained 0.1 mg of protein, 25 µl of ^{111}In with 193 Mbq of activity. This was diluted to 100 µl with saline). Each syringe contained 16 Mbq of activity after preparation, with 0.1047 g (105.4 µl), 0.1024 g (102.4 µl), and 0.1078 g (103.8 µl) being injected into the tail vein of mouse 1, 2 and 3 respectively. The last syringe underwent serial dilutions to act as a control (see Appendix 2.8). Each mouse was examined using a Bioscan nanoSPECT/CT scanner (Mediso, Budapest) (after 0 h, 3 h, 24 h and 48 h) under isofluorane anaesthesia and respiration monitoring. After scanning, the mice were culled and their organs harvested and analysed with a gamma counter for activity compared to the control.

2.5. References

1. Fan, R.; Chew, S. W.; Cheong, V. V.; Orner, B. P., Fabrication of Gold Nanoparticles Inside Unmodified Horse Spleen Apoferritin. *Small* **2010,** *6* (14), 1483-1487.
2. Aisen, P.; Listowsky, I., Iron transport and storage proteins. *Annual Review of Biochemistry* **1980,** *49*, 357-393.
3. Theil, E. C., Ferritin - structure, gene-regulation, and cellular function in animals, plants, and microorganisms. *Annual Review of Biochemistry* **1987,** *56*, 289-315.
4. Andrews, S. C., Iron storage in bacteria. *Advances in Microbial Physiology, Vol 40* **1998,** *40*, 281-351.
5. Theil, E. C., Coordinating responses to iron and oxygen stress with DNA and mRNA promoters: The ferritin story. *Biometals* **2007,** *20* (3-4), 513-521.
6. Liu, X. F.; Theil, E. C., Ferritins: Dynamic management of biological iron and oxygen chemistry. *Accounts of Chemical Research* **2005,** *38* (3), 167-175.
7. Granier, T.; Gallois, B.; Dautant, A.; Destaintot, B. L.; Precigoux, G., Comparison of the structures of the cubic and tetragonal forms of horse-spleen apoferritin. *Acta Crystallographica Section D-Biological Crystallography* **1997,** *53*, 580-587.
8. de Val, N.; Declercq, J.-P.; Lim, C. K.; Crichton, R. R., Structural analysis of haemin demetallation by L-chain apoferritins. *Journal of Inorganic Biochemistry* **2012,** *112*, 77-84.
9. Stefanini, S.; Chiancone, E.; Arosio, P.; Finazziagro, A.; Antonini, E., Structural heterogeneity and subunit composition of horse ferritins. *Biochemistry* **1982,** *21* (10), 2293-2299.
10. Uchida, M.; Klem, M. T.; Allen, M.; Suci, P.; Flenniken, M.; Gillitzer, E.; Varpness, Z.; Liepold, L. O.; Young, M.; Douglas, T., Biological containers: Protein cages as multifunctional nanoplatforms. *Advanced Materials* **2007,** *19* (8), 1025-1042.
11. Harrison, P. M.; Arosio, P., Ferritins: Molecular properties, iron storage function and cellular regulation. *Biochimica Et Biophysica Acta-Bioenergetics* **1996,** *1275* (3), 161-203.
12. Yariv, J.; Kalb, A. J.; Sperling, R.; Bauminger, E. R.; Cohen, S. G.; Ofer, S., The composition and the structure of bacterioferritin of Escherichia-coli. *Biochemical Journal* **1981,** *197* (1), 171-175.
13. Lebrun, N. E.; Andrews, S. C.; Guest, J. R.; Harrison, P. M.; Moore, G. R.; Thomson, A. J., Identification of the ferroxidase center of Escherichia-coli bacterioferritin. *Biochemical Journal* **1995,** *312*, 385-392.
14. Lawson, D. M.; Treffry, A.; Artymiuk, P. J.; Harrison, P. M.; Yewdall, S. J.; Luzzago, A.; Cesareni, G.; Levi, S.; Arosio, P., Idenitifcation of the ferroxidase active site in Ferritin. *Febs Letters* **1989,** *254* (1-2), 207-210.
15. Crow, A.; Lawson, T. L.; Lewin, A.; Moore, G. R.; Le Brun, N. E., Structural Basis for Iron Mineralization by Bacterioferritin. *Journal of the American Chemical Society* **2009,** *131* (19), 6808-6813.
16. Almiron, M.; Link, A. J.; Furlong, D.; Kolter, R., A novel DNA-binding protein with regulatory and protective roles in starved Escherichia-coli. *Genes & Development* **1992,** *6* (12B), 2646-2654.
17. Martinez, A.; Kolter, R., Protection of DNA during oxidative stress by the nonspecific DNA-binding protein Dps. *Journal of Bacteriology* **1997,** *179* (16), 5188-5194.
18. Grant, R. A.; Filman, D. J.; Finkel, S. E.; Kolter, R.; Hogle, J. M., The crystal structure of Dps, a ferritin homolog that binds and protects DNA. *Nature Structural Biology* **1998,** *5* (4), 294-303.
19. Fan, R.; Boyle, A. L.; Cheong, V. V.; Ng, S. L.; Orner, B. P., A Helix Swapping Study of Two Protein Cages. *Biochemistry* **2009,** *48* (24), 5623-5630.

20. Bevers, L. E.; Theil, E. C., Maxi- and mini-ferritins: minerals and protein nanocages. *Progress in molecular and subcellular biology* **2011,** *52*, 29-47.
21. Zhang, Y.; Fu, J.; Chee, S. Y.; Ang, E. X. W.; Orner, B. P., Rational disruption of the oligomerization of the mini-ferritin E. coli DPS through protein-protein interface mutation. *Protein Science* **2011,** *20* (11), 1907-1917.
22. Chiancone, E.; Ceci, P., The multifaceted capacity of Dps proteins to combat bacterial stress conditions: Detoxification of iron and hydrogen peroxide and DNA binding. *Biochimica Et Biophysica Acta-General Subjects* **2010,** *1800* (8), 798-805.
23. Ueno, T.; Suzuki, M.; Goto, T.; Matsumoto, T.; Nagayama, K.; Watanabe, Y., Size-selective olefin hydrogenation by a Pd nanocluster provided in an apo-ferritin cage. *Angewandte Chemie-International Edition* **2004,** *43* (19), 2527-2530.
24. Yamashita, I.; Hayashi, J.; Hara, M., Bio-template synthesis of uniform CdSe nanoparticles using cage-shaped protein, apoferritin. *Chemistry Letters* **2004,** *33* (9), 1158-1159.
25. Keyes, J. D.; Hilton, R. J.; Farrer, J.; Watt, R. K., Ferritin as a photocatalyst and scaffold for gold nanoparticle synthesis. *Journal of Nanoparticle Research* **2011,** *13* (6), 2563-2575.
26. Zhang, L.; Swift, J.; Butts, C. A.; Yerubandi, V.; Dmochowski, I. J., Structure and activity of apoferritin-stabilized gold nanoparticles. *Journal of Inorganic Biochemistry* **2007,** *101* (11-12), 1719-1729.
27. Butts, C. A.; Swift, J.; Kang, S. G.; Di Costanzo, L.; Christiansen, D. W.; Saven, J. G.; Dmochowski, I. J., Directing noble metal ion chemistry within a designed ferritin protein. *Biochemistry* **2008,** *47* (48), 12729-12739.
28. Cheung-Lau, J. C.; Liu, D.; Pulsipher, K. W.; Liu, W.; Dmochowski, I. J., Engineering a well-ordered, functional protein-gold nanoparticle assembly. *Journal of Inorganic Biochemistry* **2014,** *130*, 59-68.
29. Kramer, R. M.; Li, C.; Carter, D. C.; Stone, M. O.; Naik, R. R., Engineered protein cages for nanomaterial synthesis. *Journal of the American Chemical Society* **2004,** *126* (41), 13282-13286.
30. Galvez, N.; Sanchez, P.; Dominguez-Vera, J. M., Preparation of Cu and CuFe Prussian Blue derivative nanoparticles using the apoferritin cavity as nanoreactor. *Dalton Transactions* **2005,** (15), 2492-2494.
31. Sana, B.; Johnson, E.; Sheah, K.; Poh, C. L.; Lim, S., Iron-based ferritin nanocore as a contrast agent. *Biointerphases* **2010,** *5* (3), FA48-FA52.
32. Li, K.; Zhang, Z.-P.; Luo, M.; Yu, X.; Han, Y.; Wei, H.-P.; Cui, Z.-Q.; Zhang, X.-E., Multifunctional ferritin cage nanostructures for fluorescence and MR imaging of tumor cells. *Nanoscale* **2012,** *4* (1), 188-193.
33. Mitchell, N.; Kalber, T. L.; Cooper, M. S.; Sunassee, K.; Chalker, S. L.; Shaw, K. P.; Ordidge, K. L.; Badar, A.; Janes, S. M.; Blower, P. J.; Lythgoe, M. F.; Hailes, H. C.; Tabor, A. B., Incorporation of paramagnetic, fluorescent and PET/SPECT contrast agents into liposomes for multimodal imaging. *Biomaterials* **2013,** *34* (4), 1179-1192.
34. Lin, X.; Xie, J.; Niu, G.; Zhang, F.; Gao, H.; Yang, M.; Quan, Q.; Aronova, M. A.; Zhang, G.; Lee, S.; Leaprnan, R.; Chen, X., Chimeric Ferritin Nanocages for Multiple Function Loading and Multimodal Imaging. *Nano Letters* **2011,** *11* (2), 814-819.
35. Uchida, M.; Flenniken, M. L.; Allen, M.; Willits, D. A.; Crowley, B. E.; Brumfield, S.; Willis, A. F.; Jackiw, L.; Jutila, M.; Young, M. J.; Douglas, T., Targeting of Cancer Cells with Ferrimagnetic Ferritin Cage Nanoparticles. *Journal of the American Chemical Society* **2006,** *128* (51), 16626-16633.
36. Kitagawa, T.; Kosuge, H.; Uchida, M.; Dua, M. M.; Iida, Y.; Dalman, R. L.; Douglas, T.; McConnell, M. V., RGD-Conjugated Human Ferritin Nanoparticles for Imaging Vascular

Inflammation and Angiogenesis in Experimental Carotid and Aortic Disease. *Molecular Imaging and Biology* **2012,** *14* (3), 315-324.

37. Ardejani, M. S.; Li, N. X.; Orner, B. P., Stabilization of a Protein Nanocage through the Plugging of a Protein-Protein Interfacial Water Pocket. *Biochemistry* **2011,** *50* (19), 4029-4037.
38. Zhang, Y.; Raudah, S.; Teo, H.; Teo, G. W. S.; Fan, R.; Sun, X.; Orner, B. P., Alanine-shaving Mutagenesis to Determine Key Interfacial Residues Governing the Assembly of a Nano-cage Maxi-ferritin. *Journal of Biological Chemistry* **2010,** *285* (16), 12078-12086.
39. Bryce, C. F. A.; Crichton, R. R., Catalytic activity of horse spleen apoferritin - preliminary kinetic studies and effect of chemical modification. *Biochemical Journal* **1973,** *133* (2), 301-&.
40. Kelly, S. M.; Jess, T. J.; Price, N. C., How to study proteins by circular dichroism. *Biochimica Et Biophysica Acta-Proteins and Proteomics* **2005,** *1751* (2), 119-139.
41. Greenfield, N. J., Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols* **2006,** *1* (6), 2876-2890.

Chapter 3

Crystal structure of the hybrid maxi/mini-ferritin, Dps+E

NOTE: Parts of this chapter were undertaken in collaboration with Dr. Ralf Jauch (Genome Institute of Singapore, A*STAR, Singapore). All crystal plates were set up and the crystal structure was solved by Tom Cornell, the PhD candidate, and Dr. Jauch provided advice, expertise and arranged data collection at the Brookhaven National Synchrotron Light.

3.1. Introduction

Protein nanocages can be a source of adaptable tools and are rapidly being developed to fulfill a range of applications (see Chapter 1). However, protein folding and assembly, which are fundamental to the structure and stability of these proteins, are not fully understood and as was seen in Chapter 2, a lack of understanding of these fundamentals can hamper their development. This field often requires complex and time- and resource-intensive characterisation when establishing any application.

Expanding our understanding of the principles required for self-assembly and protein folding could enhance our ability to exploit protein nanocages and could aid in their development. While many methods already exist to assess and study proteins with complex quaternary structure (see Chapter 1), these methods are often indirect and individually often only provide limited knowledge. A technique with atomic level resolution could provide a wider and more detailed level of understanding of the entire assembled structure of a protein. This information could then be used to aid the development of novel tools such as nanoparticle formation reactors (see Chapter 2), due to the increased understanding of the structural fundamentals behind assembly. By targeting this approach toward hybrids of proteins that separately have different assembly characteristics, a fuller understanding of why they are different and how to control these differences could be discovered.

3.1.1. Aims of this investigation

As part of a project to understand the role subdomains of mini- and maxi-ferritins played in their assembly, our laboratory previously generated a number of hybrid proteins where the subdomains were swapped.[1] One of these, **Dps+E**, which was made up of Dps, a mini-ferritin, and the C-terminal E-helix of the maxi-ferritin Bfr, assembles into an oligomer

with a size intermediate between its two parents but maintains the Dps-like stoichiometry of twelve monomers. The aim of the research described in this chapter was to obtain the crystal structure of **Dps+E** and to determine why this protein assembles into an oligomer with a unique size and if the symmetry constraints of Dps, which has tetrahedral symmetry and no 4-fold axis, override those of the E-helix which is located at the 4-fold axis of symmetry in the octahedral Bfr.

3.1.2. The role of helical subdomains in mini- and maxi-ferritins, and the genesis of Dps+E

The bacterial ferritin nanocage proteins, Bacterioferritin (Bfr) and DNA binding protein from starved cells (Dps) have been extensively explored by our laboratory [2, 3, 4, 5, 6] (see Chapters 1 and 2). The monomers of Bfr and Dps, while having very different amino acid sequences, share a strikingly similar, four helix-bundle tertiary structure (Figure 3.1). This is surprising as the quaternary structures of these proteins are very different; the maxiferritin Bfr assembles into a 24-mer with octahedral symmetry and the miniferritin Dps is a 12-mer with tetrahedral symmetry. While the monomers of Bfr and Dps can be superimposed with few immediately obvious differences, the quaternary structures cannot.

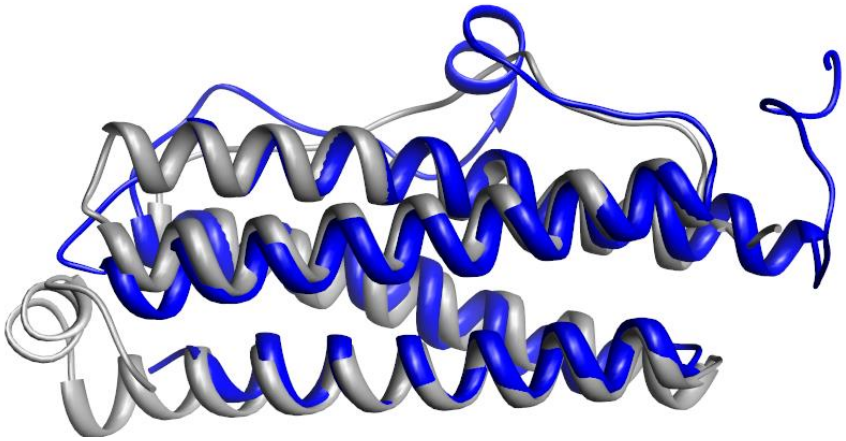


Figure 3.1 – Structural alignment of the monomer crystal structures of Bfr (grey, PDB:1BFR[7]) and Dps (blue, PDB:1DPS[8]) demonstrating the similarity of their four-helix bundle tertiary folds. Image generated with Chimera[9].

One of the differences between the tertiary structures of the two proteins is the presences of an additional, extra bundle, fifth helix in both. Bfr contains an extra helix at the C-

terminus, running 78° with respect to the bundle axis[7]. This "E-helix" is at the heart of the four-fold axis of the assembled nanocage. A four-fold axis is distinctive to octahedral symmetry and therefore it is thought that the E-helix could potentially be the driving force for the adoption of the assembled structure. Dps lacks the E-helix but instead has an additional helix, the "BC-helix", on the loop in between the second and third helices of the bundle[8]. The Dps nanocage has tetrahedral geometry which has no four-fold axis but two distinct types of three-fold axes and a two-fold axis, across which the BC-helix is positioned.

A previous study undertaken in the Orner laboratory attempted to unravel the roles of these two helixes by creating a set of mutants that incorporated all combinations of these helixes on the core four-helix bundles of both Dps and Bfr[1]. The sizes of the resulting nanocages were scrutinized using size exclusion chromatography (SEC), transmission electron microscopy (TEM) and dynamic light scattering (DLS). Through the course of this study, it was determined that the E-helix plays a larger role in assembly than does the BC-helix. In addition, several of the mutants showed no nanocage formation at all, while others displayed sizes that were no different from their parent protein. However, a protein consisting of Dps with the BC-helix and the E-helix from Bfr, "**Dps+E**", cleanly assembled into a nanocage larger than Dps, but smaller than Bfr (Figure 3.2). Sedimentation equilibrium confirmed that **Dps+E**, consistent to that of the Dps parent, was made up of twelve monomers.

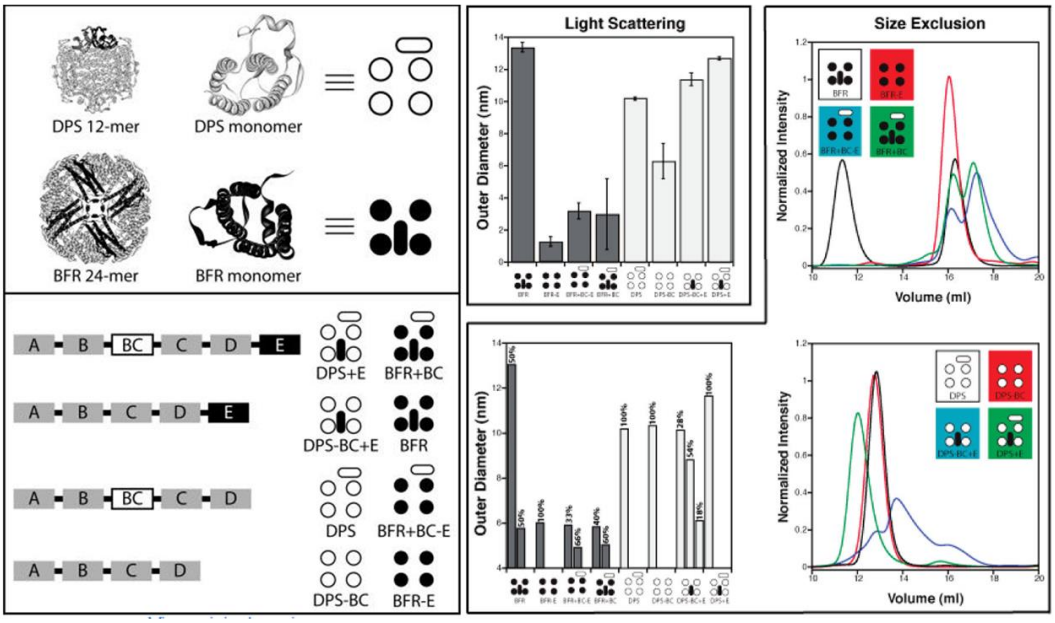


Figure 3.2 – The role of helical subdomains in mini- and maxi-ferritin assembly. (left) Schematic representation of the Dps and Bfr monomers emphasizing the four helix bundles (circles) and the BC- and E- helices (ovals). Chimera were created by deleting, adding, and swapping the BC- and E- helices between Dps and Bfr. (middle top) Hydrodynamic diameter measured by DLS. **Dps+E** is on the right. (right top) SEC analysis of Bfr derivatives, and (bottom right) Dps derivatives. **Dps+E** is the trace in green. (bottom middle) Quantitative analysis of SEC data. This figure is from Orner and co-workers (2009)[1].

Although it had been established that **Dps+E** was larger than Dps but smaller than Bfr and was made up of twelve monomers like Dps, it was unclear how this actually manifested in the structure of **Dps+E**. Based on this data, we had two initial questions. First, because the E-helix in Bfr is positioned at the four-fold axis of symmetry in octahedral Bfr, but Dps is tetrahedral, without any four-fold symmetry, we were curious if the a four-fold or three-fold symmetric quaternary structure would prevail. In other words, would the structural preference of Dps or the Bfr E-helix win? The second question involved an older paper from the ferritin literature[10]. This paper speculated that the E-helix, which is normally pointing into the protein nanocage cavity in octahedral maxi-ferritins, could “flop” in and “flip” out of the nanocage in a dynamic manner. We therefore were curious whether **Dps+**E could be frozen in a “flipped out” state. (Figure 3.3). The best way to obtain answers to these questions is to obtain the crystal structure of **Dps+E**. A crystal structure should be able to definitively answer these questions and provide high resolution information about key protein-protein interactions and how they

have been altered from those in Bfr or Dps. In addition the crystal structure could provide insight into the structural energetics of nanocage structures and this insight could help to engineer similar assemblies in the future.

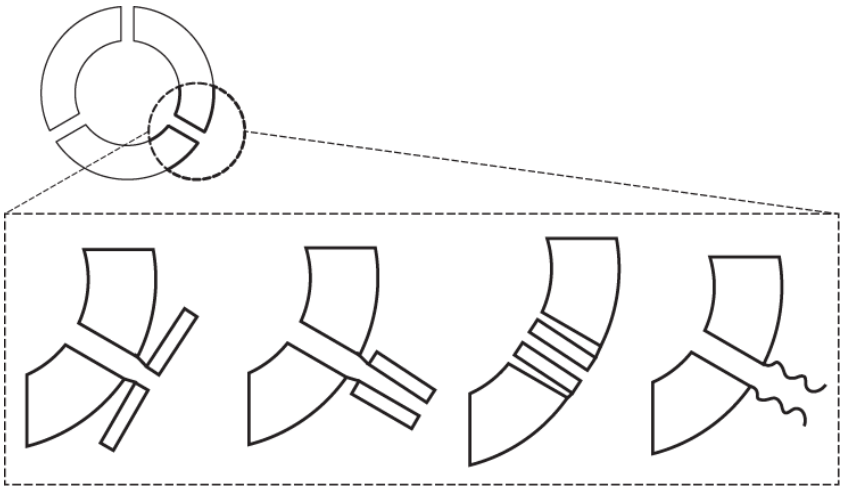


Figure 3.3 – Schematic representation of the possible presentation of the E-helix residues from Bfr (rectangle) in the hybrid protein **Dps+E** (assembled circle). Establishing the position of these residues is one of the aims of determining the crystal structure of the protein as described in this chapter.

3.2. Results and discussions

3.2.1. Cloning, expression and purification of Dps+E

The gene encoding **Dps+E** was transferred to an expression plasmid (pET-46, Novagen) that provides a fusion with a small His_6 tag on the N-terminus, thus avoiding extended purification methods requiring the removal of large purification tags with protease digestion as were used in Chapter 2. In addition, this new purification protocol bypassed problems with fragmentation of **Dps+E** initially observed in proteolytic conditions (see Appendix 3.3 for SDS-PAGE). Expression and purification and characterisation of the protein confirmed that it was intact and assembled as a 12-mer (see Appendices 3.2, 3.3, 3.4, 3.5 and 3.6 for the full amino acid sequence, SDS PAGE, TEM, mass spec and SEC). In addition, SEC analysis of this protein with the His_6 tag resulted in a similar elution profile to that previously reported (Figure 3.2) further confirming that its unique conformation was robust.

3.2.2. Crystallisation of Dps+E

Although there are many existing ferritin crystal structures in the protein data bank (PDB), their crystallisation conditions are highly varied. Therefore, the screening for crystallisation preceded without focusing on previously used conditions. Using this purified protein, high throughput screens were initially constructed using the JCSG+, Classic (Qiagen) and Morpheus (Molecular dynamics) suites to find first generation crystallisation conditions. The most promising conditions contained polyethylene glycol (PEG) which lead us to optimise the conditions with screens using the PEG I and PEG II (Qiagen) suites. These results were the basis of further optimisation (see Appendix 3.7 for full range of conditions screened). It should be noted that the high throughput screens were set up robotically using the sitting drop method, while all other conditions were set up by hand using hanging drop (Figure 3.4).

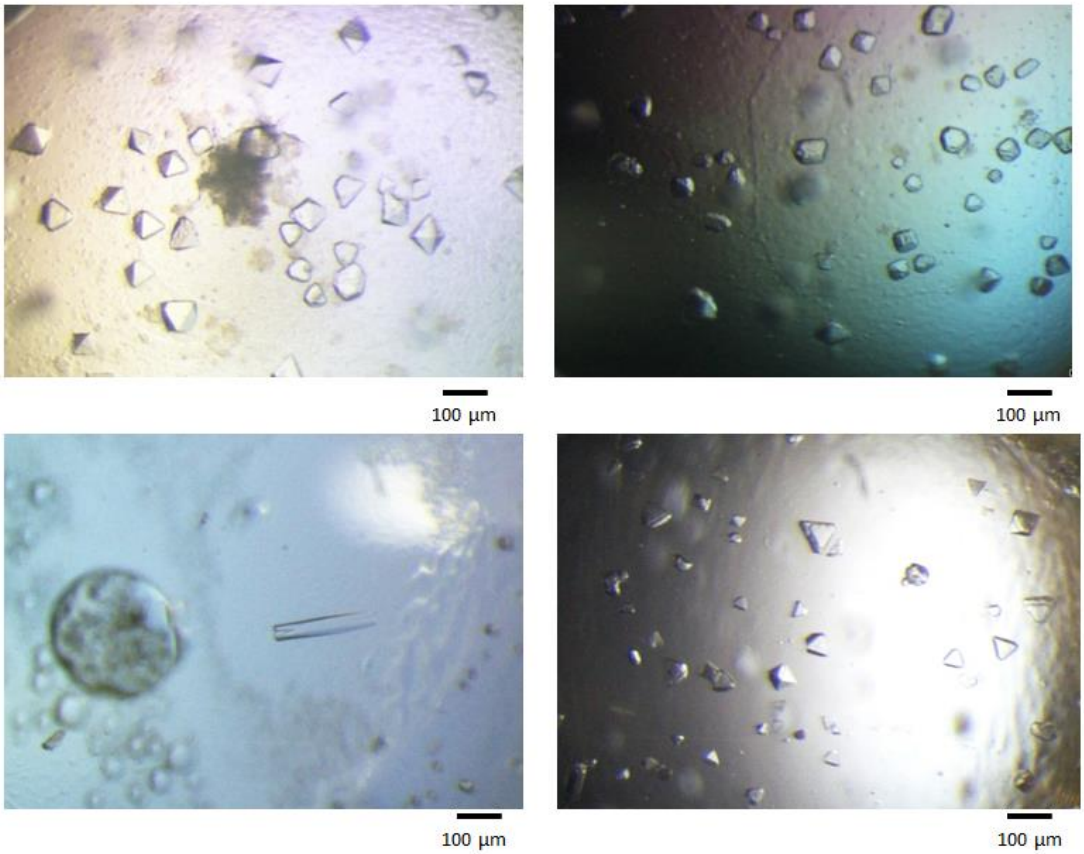


Figure 3.4 – Preliminary crystallization conditions of **Dps+E** obtained through screening. (top left) 16% PEG3350, 0.2 M magnesium chloride, 0.1M Tris·HCl pH 8.6 (top right) 16% PEG3350, 0.2 M magnesium formate, 0.1M Tris·HCl pH 8.6 (bottom left) 18% PEG3350, 0.1 M trisodium citrate, 0.1M Tris·HCl pH 8.6 (bottom right) 12% PEG6000, 0.2 M lithium sulfate, 0.1M Tris·HCl pH 8.6. All conditions in this figure used a protein concentration of 15 mg/ml.

The majority of conditions that produced crystals contained PEG 3350 and magnesium chloride. However, when inspected for preliminary diffraction patterns, none of the crystals grown in the presence of magnesium chloride diffracted. But Diffraction data however, was obtained from crystals grown in the presence of lithium sulfate. Therefore these conditions were the basis of further optimisation to increase crystal size (Figure 3.5).

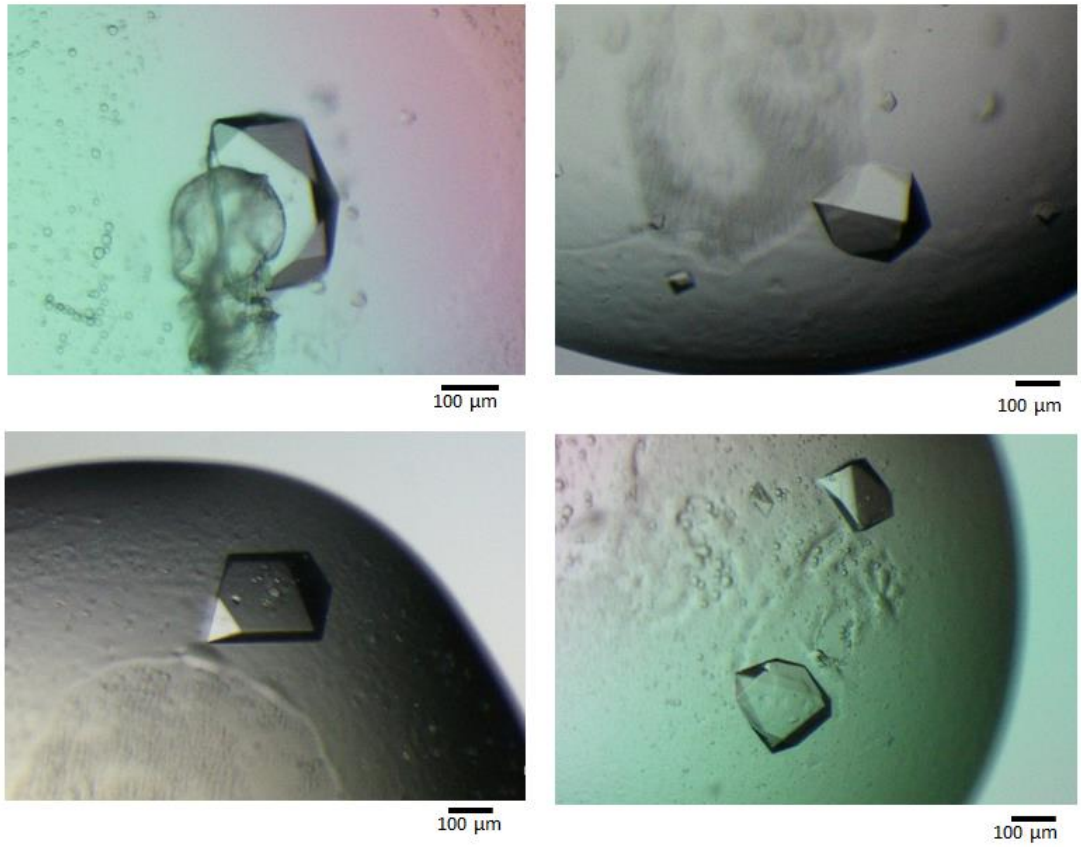


Figure 3.5 – Optimised crystallization conditions for **Dps+E** (top left) 16% PEG1000, 0.1 M lithium sulfate, 0.1 M sodium acetate trihydrate pH 4. (top right) 18% PEG1000, 0.2 M lithium sulfate, 0.1M sodium acetate trihydrate pH 4.2. (bottom left) 20% PEG1000, 0.3 M lithium sulfate, 0.1 M Sodium acetate trihydrate pH 4.2. (bottom right) 18% PEG1000, 0.1 M lithium sulfate, 0.1 M sodium acetate trihydrate pH 4.8. All conditions in this figure used a protein concentration of 15 mg/ml.

Several of these crystals were sent to the Brookhaven National Synchrotron Light Source for analysis. Crystals grown in 18% PEG1000, 0.2 M lithium sulfate, 100 mM sodium acetate trihydrate, pH 4.2 provided the highest resolution diffraction data.

Initial attempts to solve the crystal structure of **Dps+E** were successful and resulted in a 2.0 Å structure of assembled **Dps+E**. A polyalanine model of the 1DPS crystal structure monomer (obtained with Chainsaw[11] software) was used for molecular replacement analysis with Phaser[12] software. As it was already determined through sedimentation equilibrium that **Dps+E** was made up of twelve monomers, molecular replacement was performed using twelve polyalanine monomers. After molecular replacement, the structure was built and refined using

Coot[13] and Refmac[14] and validated with Phenix[15], which provided the refined crystal structure of **Dps+E** (Table 3.1).

Dps+E

[a]Values for the highest resolution shell in parentheses.

Method – Vapour diffusion, hanging drop
pH – 4.2
Temperature – 292.0 K
Method – 18% PEG 1000, 0.1 M sodium acetate trihydrate, 0.2 M Lithium sulfate

Space group	P 21 21 21
Cell dimensions	
	A = 103.14 Å
	B = 104.72Å
	C = 207.76Å
	α = 90.0 º
	β = 90.0 º
	γ = 90.0 º
Resolution (Å)	50.00 – 2.00 (2.07-2.00)
Rmerge (%)	6.9 (51.7)
Completeness (%)	93.5 (98.5)
I/σI	
Redundancy	27.8 (4.4)
Refinement	12.8 (11.9)
Resolution (Å)	46.8 – 1.99
No. Reflections used	135612
Rwork/Rfree (%)	19.1/23.8
No. Atoms	
Protein	16696
Water	1865
Cl^-	2
Non hydrogen atoms	16696
Mean isotropic B value	24.433
R.M.S deviations from ideal	
Bond lengths (Å)	0.012
Bond angles (◦)	1.494
Ramachandran analysis (%)	

Favoured	98.3
Additionally allowed	17
Disallowed	0
R-Factor (All)	0.1895
R-Factor (Observed)	0.1895
R-work	0.1868
R-free	0.2408
R-free selection details	Random

Table 3.1 – Data collection and refinement statistics for the crystal structure of **Dps+E**

3.2.3. Analysis of the Dps+E crystal structure

Initial inspection of the **Dps+E** crystal structure showed that it assembles into a very similar tetrahedral symmetric protein nanocage to that of the parent protein, Dps (Figure 3.6). The global Route Mean Square Deviation (RMSD) for the overlaid monomers with identical sequences was 0.185 Å demonstrating strong similarity (for individual amino acid RMSD comparisons see Appendix 3.8). Thus, one question posed at the outset of this project is definitively answered. The symmetry requirements of the parent protein, Dps, overcome those of the residues from the Bfr E-helix. The hybrid protein, **Dps+E** clearly assembles into a protein nanocage that is tetrahedral and not octahedral, and the residues from the E-helix are positioned at a three-fold axis of symmetry, not a four-fold symmetry axis as they are in Bfr.

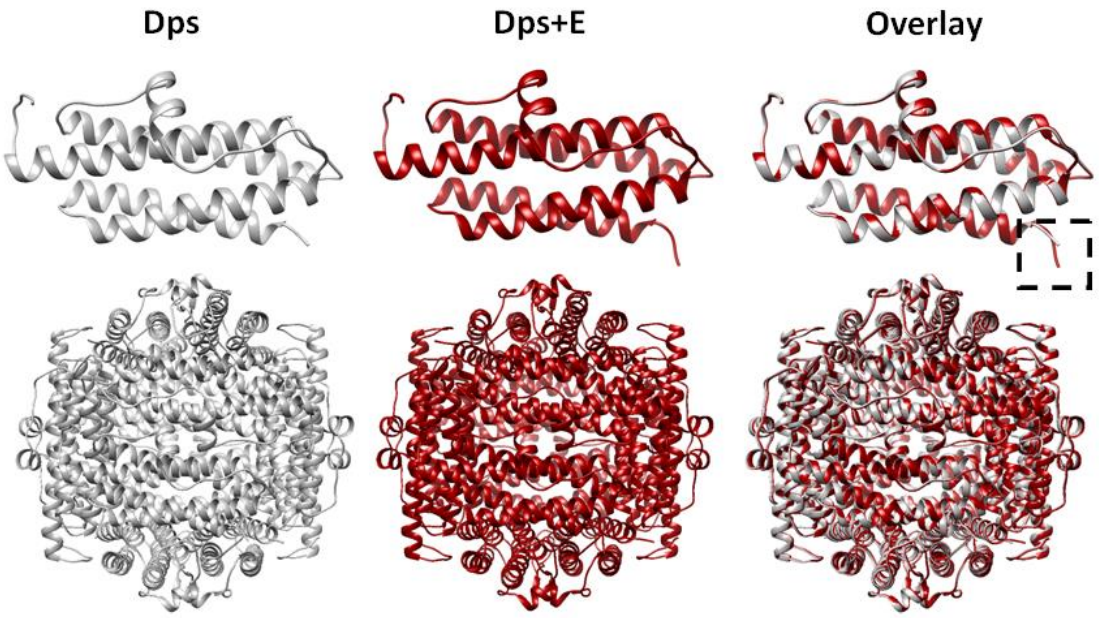


Figure 3.6 – Comparison of the crystal structures for wild type Dps (PDB:1DPS, grey) and for **Dps+E** obtained in research described in this chapter (red). Comparison of monomers (top) and assembled nanocages (bottom). Highlighted region (black dashed box) shows the position of the C-terminus for both structures. Note that the electron density for all the residues from the Bfr E-helix is poor, although it is clear that they are flipped out of the nanocage, the exact structure of this entire domain is undetermined. Image created with Chimera[9] with overlay generated with the 'match' function.

The second question this research posed can also be definitively answered, however not in as satisfying way. Unfortunately, our crystal structure of **Dps+E** had only limited electron density in the C-terminal domain made up of the residues from the Bfr E-helix. Although several residues from the fusion could be modeled, the structure of the entire domain is undetermined, however, enough data could be obtained to confirm that these residues are pointing external to the assembled protein and, thus, this domain is "flipped out" of the nanocage (Figure 3.7). Taken together, these results explain how the **Dps+E** nanocage can be made up of only twelve monomers, as confirmed by sedimentation equilibrium, but appear larger than Dps and smaller than Bfr in SEC and DLS; the E-domains fused to the C-terminus of every monomer and flipped out of the 12-mer cage results in a particle of larger hydrodynamic radius.

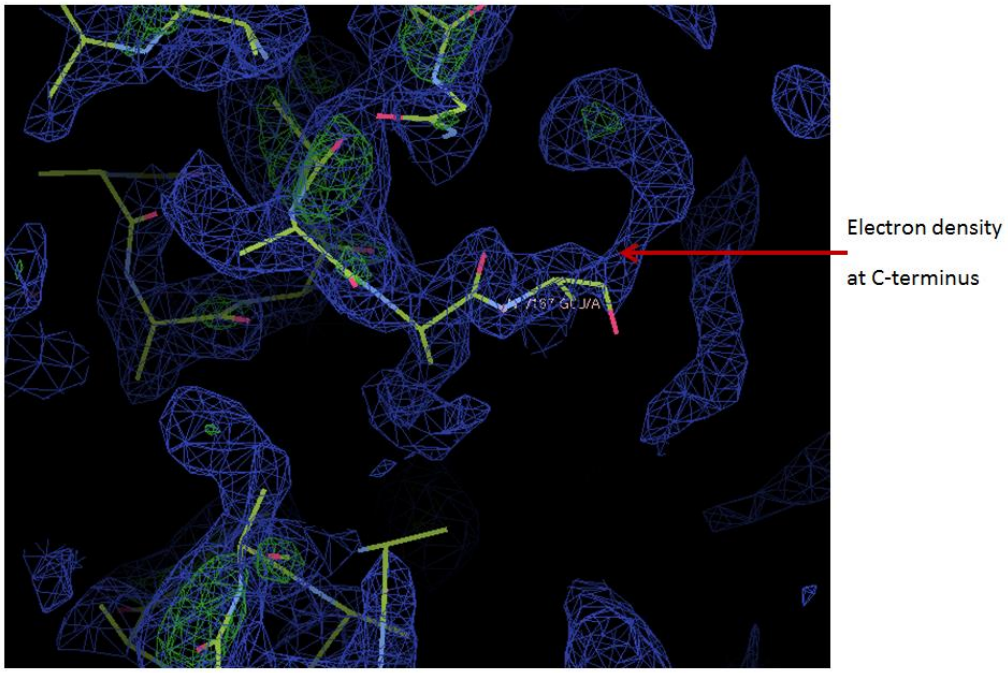


Figure 3.7 – Electron density map with constructed structure for the protein **Dps+E** highlighting the fourth amino of the fused E-helix domain of Bfr. Note the fall off in electron density beyond this residue. Image was generated with Coot software[13].

While the aims set out in this project, (to obtain the DPS+E crystal structure, confirm the overall symmetry, determine the position of the E-domains with respect to the nanocage, and understand why the protein is larger in size than Dps) were all fulfilled, it was disappointing to not be able to observe the entire E-domain. With these termini projecting away from the nanocage on the three-fold axes of symmetry, it was thought that they might have engineering uses as “handles” for applications like drug delivery or supra-assembled materials. Therefore we intend to pursue high resolution electron microscopy studies to understand the relative positions of these domains

3.3. Conclusions

The hybrid ferritin protein **Dps+E**, which had previously been shown to be a 12-mer but larger in size than the miniferritin Dps, was crystalized and its high resolution crystal structure was obtained. Examination of this crystal structure showed that the tetrahedral symmetry of Dps, was retained in the assembled **Dps+E** protein nanocage and that the addition of the E-domain from octahedral Bfr was not enough to change the symmetry. Moreover, it was determined that the E-domain was forced into a "flipped out" state where it projects away from the nanocage. However, as the electron density at this domain was poorly defined, the entire structure of the domain could not be fully determined. It is presumed that the projection of twelve copies of this domain away from the nanocage is the reason the protein was determined to be larger in size than Dps.

While answering fundamental questions, a unique high resolution protein nanocage structure was established. However, obtaining a crystal structure is time and resource intensive, and it gives little information about the thermodynamics and mechanism of assembly. Additionally it is also not immediately clear how the information provided by this structure could dramatically aid in the design of other novel quaternary structures or help to overcome the limitations in protein nanocage stability at different conditions as was experienced in the project described in Chapter 2. Clearly it is necessary to develop new techniques that are complementary to existing ones, both to study the fundamentals of quaternary structure and to help engineer new structures in a rapid and dynamic manner. The next three chapters of this thesis will present our work in this direction.

3.4. Methods

3.4.1. Cloning, protein purification and characterisation

3.4.1.1. Construction of the DPS+E plasmid

The Dps+E gene was prepared for ligation to the plasmid by first extending the ends to include LIC nucleotide sequences. This was done in a mixture containing reaction buffer (Fermentas, 5 µl of 10xPfu buffer), dNTP mix (Fermentas, 2 µl of a solution containing dATP, dTTP, dGTP, and dCTP at 2 mM each), forward oligonucleotide primer (400 ng, Suprenom, Singapore, see Appendix 3.1), reverse oligonucleotide primer (400 ng, Suprenom, Singapore, see Appendix 3.1 for primers used), the dsDNA template of Dps+E (100 ng) and Pfu polymerase (Fermentas, 2 µl of 2.5 U/µl) and made up to a total volume of 50 µl with deionised water. The PCR conditions involved an initial step at 95 °C for 30 s, followed by 30 cycles of amplification (95 °C for 30 s, 55 °C for 45 s and 72 °C for 1 min) followed by incubation (72 °C for 7 min). The resulting PCR product was isolated by gel excision from a 1% agarose gel. The gene was then treated with T4 polymerase in order to create the required complementary overhangs. This was done in a reaction mixture containing NEB reaction buffer 2 (NEB, 2 µl of 10x buffer), dATP (NEB, 2 µl of 25 mM), DTT (Sigma, 1 µl of 100 mM), BSA (NEB, 0.2 µl of 100x buffer), Dps+E extended gene (0.3 pmol) and T4 DNA polymerase (NEB, 0.6 µl of 5 U/µl) and made up to a total volume of 20 µl with deionised water. This reaction was incubated (30 min at 22 °C) followed by heat inactivation (75 °C for 20 min). The insert was annealed to plasmid pET-46 (Novagen) by adding insert (1 µl) to the plasmid (0.5 µl) and incubating the solution (30 min, room temp). A solution of EDTA (1 µl of 100 mM) was then added and the mixture was further incubated (30 min). This solution (2.5 µl) was then electroporated into electrocompetent Novablue *E.coli* cells. The resulting colonies were assessed by colony PCR using a reaction mixture containing Taq reaction buffer (Fermentas, 2.5 µl of 10x buffer), $MgCl_2$ (Fermentas, 1.75 µl of 25 mM) dNTP mix (Fermentas, 2 µl of a solution containing dATP, dTTP, dGTP, and

dCTP at 2 mM each), T7 promoter primer (150 ng, Suprenom, Singapore), T7 terminator primer (150 ng, Suprenom, Singapore), Taq DNA polymerase (Fermentas, 0.25 µl of 5 U/µl), colony suspension (10 µl of a 50 µl total colony suspension in deionised water) and 5 µl of deionised water. The PCR conditions involved an initial step of 95 °C for 10 min, followed by 30 cycles of, 95 °C for 30 s, 59 °C for 30 s and 72 °C for 30 s, followed by 72 °C for 7 min. All samples were visualised using a 1% agarose to confirm band size. Any colonies showing the correct band size in the colony PCR were grown then isolated using miniprep (Invitrogen) and sequenced (see Sequence Appendix 3 for full gene sequence and Appendix 3.1 for primers used).

3.4.1.2. Protein production and purification

The vector containing the correct **Dps+E** gene was electroporated into BL21 *E.coli* cells and plated. A colony was taken and placed into LB and incubated (5 ml, 37 °C, overnight) to generate a pre-culture. These cells were then grown in LB (500ml inoculated with 50 µl/ml of carbenicillin, 37 °C). Protein expression was then induced at an OD_{600} 0.6 by the addition of IPTG (400 mM) and incubated for an additional 3 hours at 20 °C. The cells were isolated by centrifugation (10,000 rpm, 20 min at 4 °C). The cell pellet was resuspended with lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 40 mM Imidazole, pH 8) and sonicated (15 min, with 10 s off and on pulses, Sonics vibra cell sonicator). The protein solution was clarified by centrifugation (10,000 rpm, 20 min at 4 °C) and then filtered (Pall, 0.2 µm). His-tagged proteins were isolated from cell lysates via His trap chromatography. The filtered lysate was injected onto a GE 5ml Histrap FF Column using wash buffer (30 mM NaH_2PO_4, 300 mM NaCl, 40 mM Imidazole, pH 7.6) at 1 ml/min. The protein of interest was eluted from the His trap using elution buffer (30 mM $NaHPO_4$, 300 mM NaCl, 500 mM Imidazole, pH 7.6). The partially purified protein sample was bound to a GE HiTrap Q HP ion exchange column using IEX binding buffer (50 mM Tris.HCl, 10 mM NaCl, pH 7.6) at 2 ml/min. An elution gradient using IEX elution buffer (50 mM Tris.HCl, 500 mM NaCl, pH 7.6) was established with the protein of interest eluting between 26-30% elution buffer. The last step of purification was to subject the partially pure sample onto a GE

Hiload16/60 Superdex 200pg 120 ml size exclusion column (5 ml injection, running buffer - 50 mM NaH_2PO_4, 50 mM NaCl, pH 7.0) at 1 ml/min. The extent of protein purification was assessed through SDS-PAGE (see Appendix 3.3 for SDS-PAGE).

3.4.2. Crystallisation conditions

3.4.2.1. Crystallisation

Purified **Dps+E** protein underwent buffer exchange and was concentrated (15 mg/ml, 10 mM NaCl, 10 mM HEPES, pH 7) for crystal screening. Initial high throughput screens were set up using robotics in 96 well plates with a sitting drop technique (JCSG+ (Qiagen), Classic suite (Qiagen), Morpheus suite (molecular dynamics), PEG suite I (Qiagen) and PEG suite II (Qiagen)). Further crystal screening took place in 15 well plates using the hanging drop technique (see Appendix 3.7 for all crystal screening conditions). The crystal that gave the diffraction pattern that was solved was grown using 0.2 M Lithium sulfate, 18% PEG 1000, pH 4.2. Crystals grew after 3 - 4 days at 19 °C.

3.4.2.2. Data collection, processing and structure solution

A 1.99 Å data set was obtained from the National Synchrotron Light source (NSLS) at Brookhaven. The data was integrated, scaled and merged using HKL2000[16]. PHASER[12] (CCPi4) was used for molecular replacement experiments, using 12 copies of a poly-alanine monomer of wild type Dps from *E.coli* (1DPS). This structure has around an 85% similarity with the expected new structure. PARROT was used to improve the phases from the molecular replacement. The model was built manually in COOT[13] using 2Fo-Fo and Fo-Fc maps. The refinement was done using REFMAC5[14] from CCP4i applying NCS restraints and a weighting term of 0.2. Translation/Liberation/Screw (TLS) refinement was conducted towards the final steps of the refinement[17]. Each monomer of the protein nanocage was assigned its own TLS group, giving 12 TLS groups in total. Final validation was done using PHENIX.VALIDATE[15].

3.4.2.3. Calculation of RMSD

The coordinates for the crystal structures of both wild type Dps (PDB:1DPS[8]) and **Dps+E** were aligned using Chimera software[9] with the Matchmaker tool, which uses a Needleman-Wunsch alignment algorithm. This tool focuses on aligning alpha carbon atoms after sequence alignments have been performed. From this overlay, least square fitting is used to calculate RMSD.

3.5. References

1. Fan, R.; Boyle, A. L.; Cheong, V. V.; Ng, S. L.; Orner, B. P., A Helix Swapping Study of Two Protein Cages. *Biochemistry* **2009,** *48* (24), 5623-5630.
2. Fan, R.; Chew, S. W.; Cheong, V. V.; Orner, B. P., Fabrication of Gold Nanoparticles Inside Unmodified Horse Spleen Apoferritin. *Small* **2010,** *6* (14), 1483-1487.
3. Ardejani, M. S.; Li, N. X.; Orner, B. P., Stabilization of a Protein Nanocage through the Plugging of a Protein-Protein Interfacial Water Pocket. *Biochemistry* **2011,** *50* (19), 4029-4037.
4. Zhang, Y.; Fu, J.; Chee, S. Y.; Ang, E. X. W.; Orner, B. P., Rational disruption of the oligomerization of the mini-ferritin E. coli DPS through protein-protein interface mutation. *Protein Science* **2011,** *20* (11), 1907-1917.
5. Zhang, Y.; Orner, B. P., Self-Assembly in the Ferritin Nano-Cage Protein Superfamily. *International Journal of Molecular Sciences* **2011,** *12* (8), 5406-5421.
6. Ardejani, M. S.; Chok, X. L.; Foo, C. J.; Orner, B. P., Complete shift of ferritin oligomerization toward nanocage assembly via engineered protein-protein interactions. *Chemical Communications* **2013,** *49* (34), 3528-3530.
7. Dautant, A.; Meyer, J. B.; Yariv, J.; Precigoux, G.; Sweet, R. M.; Kalb, A. J.; Frolow, F., Structure of a monoclinic crystal form of cytochrome b1 (bacterioferritin) from E-coli. *Acta Crystallographica Section D-Biological Crystallography* **1998,** *54*, 16-24.
8. Grant, R. A.; Filman, D. J.; Finkel, S. E.; Kolter, R.; Hogle, J. M., The crystal structure of Dps, a ferritin homolog that binds and protects DNA. *Nature Structural Biology* **1998,** *5* (4), 294-303.
9. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **2004,** *25* (13), 1605-1612.
10. Luzzago, A.; Cesareni, G., Isolation of point mutations that affect the folding of the H-chain of human ferritin in Escherichia-coli. *Embo Journal* **1989,** *8* (2), 569-576.
11. Stein, N., CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *Journal of Applied Crystallography* **2008,** *41*, 641-643.
12. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., Phaser crystallographic software. *Journal of Applied Crystallography* **2007,** *40*, 658-674.
13. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta Crystallographica Section D-Biological Crystallography* **2010,** *66*, 486-501.
14. Murshudov, G. N.; Vagin, A. A.; Dodson, E. J., Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D-Biological Crystallography* **1997,** *53*, 240-255.
15. Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D-Biological Crystallography* **2010,** *66*, 213-221.
16. Otwinowski, Z.; Minor, W., Processing of X-ray diffraction data collected in oscillation mode. *Macromolecular Crystallography, Pt A* **1997,** *276*, 307-326.
17. Painter, J.; Merritt, E. A., Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallographica Section D-Biological Crystallography* **2006,** *62*, 439-450.

Chapter 4

Direct detection of maxi-ferritin assembly with the fluorescent probe FlAsH-EDT_2

NOTE: Parts of this chapter were taken from published work with the candidate as first author (see appendix – Published Work 1)[1].

4.1. Introduction

The exploitation of protein nanocages for novel applications such as encapsulated nanoparticle genesis can be impeded by inherent challenges in working with these proteins. It was seen in Chapter 2 that one of these difficulties is protein instability. Furthermore, these complex protein assemblies can be difficult to purify, although our laboratory has extensive experience in this regard[2, 3, 4, 5]. Adding to the challenges, these proteins, upon purification, require extensive biophysical characterisation to determine their assembly state, but the techniques that are commonly employed for this characterisation are often not directly related to nanocage formation.

The time it takes to overcome these challenges can delay innovation in this field. In addition, the lack of understanding into the fundamentals of how protein quaternary structure forms, and why it fails under different conditions inhibits the ability to troubleshoot and resolve problems. Protein crystallography (see Chapter 3) can provide some degree of insight into what residues are involved in protein-protein interactions and can even be used to estimate binding thermodynamics, but it is low throughput and requires conditions that are very dissimilar from those used for the generation of nanoparticles. In addition crystallography provides little information into the mechanisms of folding, unfolding, assembly, disassembly or amorphous aggregation. Furthermore, because it is a fundamental technique by nature, even if it did provide explanations as to why a particular protein is problematic to a particular application, it would provide no direct solutions to the problem.

The development of a tool for the rapid and direct detection of specific protein nanocage oligomerization states would have great utility in this field by helping to overcome some of the protein-specific problems that challenge the field. It could provide information on specific oligomerization state formation in different conditions, such as those used for the

production of nanoparticles. If it were rapid enough it might shed light on the dynamics and mechanisms of assembly. Moreover, if this technique were applicable to living cells, it could be expanded to a high throughput protein library screen to detect protein nanocages with novel properties using a “one clone/one bacterium” cell sorting strategy.

Such an assay must be robust on multiple levels. It must be able to identify both the protein of interest over any background proteins and a specific oligomerization state of the protein of interest over any other oligomerization states. Additionally, it should implement a flexible design so as to be exportable to mutants of a specific protein nanocage or other protein nanocage systems in general. Moreover, it should be able to scrutinize a range of conditions for the evaluation of protein quaternary structure in different environments. Finally, it should behave well in complex solutions, a criterion which would allow the bypass of the need to purify and fully characterise the proteins and thus permit a substantial increase in throughput.

4.1.1. Aims of this investigation

Many high resolution protein nanocage structures are available (PDB: 2Y3Q, 1BFR, 1BCF), and although these structures have paved the way for rational engineering and design of protein nanocages with application-bespoke properties, rational design can be time-consuming and resource intensive due to the necessity of iteratively purifying each mutant followed by extensive biophysical characterisation with techniques that are often not directly related to nanocage formation. The investigation described in this chapter aims to develop a novel method for the direct detection of nanocage oligomerization with specificity of the nanocage over other states. This technique could be used to rapidly ascertain the oligomerization state of various nanocage proteins or to screen protein nanocage stability in varying conditions without extended biophysical analysis. This assay could also be expanded to

the screening of libraries of mutant proteins for their ability to form more stable nanocage structures.

4.1.2. Bacterioferritin properties to be exploited for the assembly assay

As described in chapter 2, *E. coli* Bacterioferritin (Bfr) is a maxi-ferritin which buffers cellular iron by storing it as a mineral inside the hollow cavity of the assembled state[6, 7, 8]. Bfr assembles from twenty-four identical monomers into an octahedral nanocage with two-, three- and four-fold symmetric protein-protein interfaces[9]. The Bfr protein nanocage is a useful starting point in any new investigation as it is extremely well characterised and has been extensively employed in many laboratories including the Orner laboratory[2, 3, 10]. Previous studies on the self-assembly of Bfr and ferritins in general have shown that monomers rarely persist in solution on their own unless severely mutated, and, for most ferritins, the dimer is the most prevalent intermediate observed[2, 3]. Bfr exists *in vitro* as a mixture of the nanocage form and the anti-parallel dimer in a near equal ratio resulting in a unique challenge for a detection technique that can be specific for a single oligomerization state[4]. In the Bfr dimer, the termini diverge whereas in the nanocage form they converge, it is thought that this could direct the placement of the cysteine pairs in the design of a FlAsH binding site that is specific for the nanocage state. The fact that Bfr forms a mixture of states can help in the development of this technique for the determination of the specificity of early generation designs.

4.1.2.1. Control proteins: Mutations that influence the ratio of oligomerization states of Bfr

The development of a technique to detect a specific oligomerization state of Bfr will require control proteins that have altered nanocage to dimer ratios. To date there is no rapid and direct method for determining the oligomerization states of nanocages other than expressing, purifying and characterizing them with mostly low throughput biophysical techniques such as size exclusion chromatography (SEC), dynamic light scattering (DLS) and transmission electron microscopy (TEM). Thus, minimal work has been undertaken to

manipulate the oligomerization state of ferritins. Our laboratory has investigated the quaternary structure of Bfr by disrupting the interaction network through alanine scanning[11] to discover 'hot spot' residues along key protein-protein interactions [3, 12, 13]. Additionally we have used computational approaches to more efficiently pack the protein-protein interfaces by filling water pockets with bulky hydrophobic residues[4, 10]. Through these studies, we have discovered a number of mutants with distinct ratios of nanocage to dimer. Some of these mutants that are relevant to the study described in this chapter are N23F, which exists as 70% nanocage, D118F, which favours the nanocage to such an extent that almost no dimer is observable. These works demonstrate that Bfr is well studied, and when combined with the availability of the well-characterised mutants to act as controls, can provide a rich system to act as a proof of concept for the development of the FlAsH-based technique.

4.1.3. FlAsH technology for the detection of protein-protein interactions

As described in Chapter 1 there are many existing techniques that can be used for the detection of protein-protein interactions which is the goal of this investigation. However these techniques are not without theirs flaws, often requiring the fusion and creation of large protein constructs which can inhibit native properties and can limit expansion and throughput of any developed screen.

A new probe called, fluorescein arsenic hairpin binder (FlAsH, Figure 4.1 –A), described by Tsien and co-workers only fluoresces once it is bound to four correctly spaced sulfur atoms presented on cysteine amino acids along the face of a α-helix (Figure 4.1- B)[15]. FlAsH technology works by only requiring the addition or alteration of four amino acids to cysteines and can bypass problems seen with larger fusions. Development of FlAsH technology showed that binding to the hexapeptide sequence CCPGCC increased observed fluorescence signal (Figure 4.1 – C)[16]. FlAsH had being used to monitor several protein attributes such as location[17] or aggregation rate[18], but Schepartz and co-workers expanded this to be able to detect

protein-protein interactions using this small cell permeable probe. They showed that the hexapeptide sequence could be split into two bipartite spilt cysteine tags that still allowed for FlAsH binding sites either when placed on termini of the same protein that appear close in space (Figure 4.1 – D) or on two interacting peptides (Figure 4.1 – E)[19]. This technique, known as bipartite tetracysteine display, could provide a better approach for the detection of protein-protein interactions seen on nanocages as it only needs the addition of four amino acids, and allows for greater control via the external addition of the probe.

By designing cysteine rich FlAsH-EDT_2 binding sites at interfaces that only appear upon nanocage formation, it is hoped that a rapid and direct assay for self-assembly can be found (Figure 4.1 – F). A system that can identify specific oligomerization states with increased throughput and with respect to states in cell lysates or *in vivo,* would greatly advance research on protein nanocages and protein self-assembly in general.

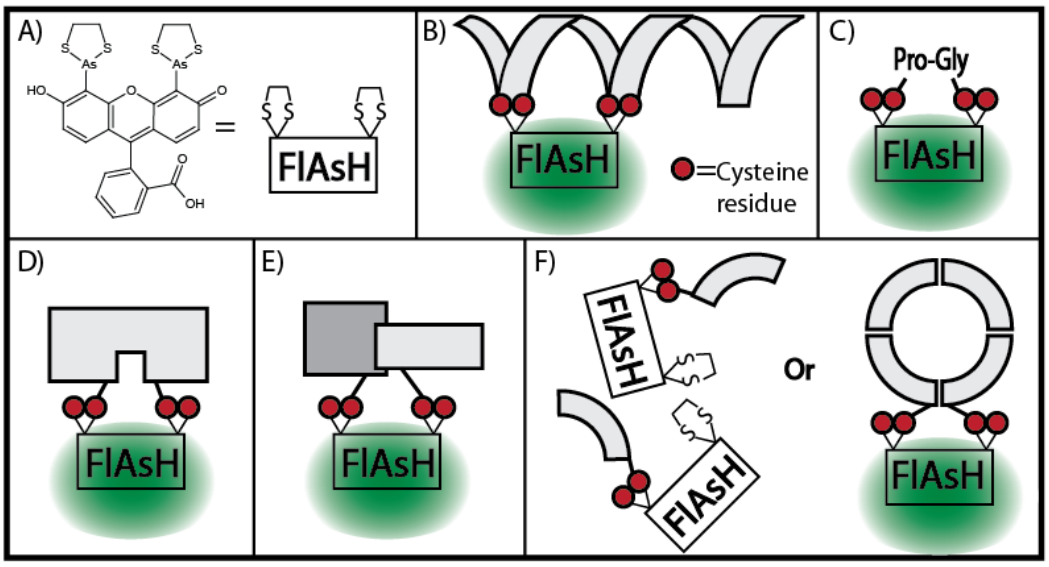


Figure 4.1 – Conceptual evolution of FlAsH binding site design. (A) Structure and schematic of FlAsH-EDT_2. (B) Initial presentation of four cysteines on one face of a α-helix. (C) Optimized hairpin peptide. (D) Bipartite cysteine display with the two cysteine pairs placed proximal on each termini of the same protein. (E) Bipartite cysteine display with cysteine pairs straddling a protein-protein interface between interacting proteins. (F) Strategy to detect self-assembly of nanocage structures[1].

4.2. Results and discussion

4.2.1. Design of Bfr controls and optimisation of experimental parameters

To evaluate our reagents and initial experimental set up it was necessary to establish robust positive and negative Bfr controls. A positive control was constructed by placing the ideal FlAsH-EDT_2 binding site, the peptide CCPGCC (see above), onto the N-terminus of Bfr (**BfrCCPGCC-N-term**) while the wild-type protein acted as a negative control (**BfrWT**). The positive control is oligomerization state independent, as every monomer with this sequence should be able to bind to FlAsH-EDT_2 no matter the assembly state whereas the negative control has no binding site.

Initial optimisation of conditions was performed in bacterial lysates so as to not only speed the process by removing the need to purify proteins, but also to bring the system one step closer to examining self-assembly in *in vivo* conditions. Both of the controls described above, were analysed using a fluorimeter for lysate samples with different concentrations of FlAsH-EDT_2 incubated for two hours (Figure 4.2). As expected, the signal from the positive control increased significantly as FlAsH-EDT_2 concentrations were increased, while the negative control remained low, demonstrating that the controls, reagents and instrumentation behaved as expected and that a dose dependent signal could be achieved in lysate conditions for **BfrCCPGCC-N-term**. While there was a clear significant difference between **BfrCCPGCC-N-term** and **BfrWT** across different FlAsH-EDT_2 concentrations, the widest window was at the highest concentration. Thus, 1 μM was used in all subsequent experiments unless stated otherwise

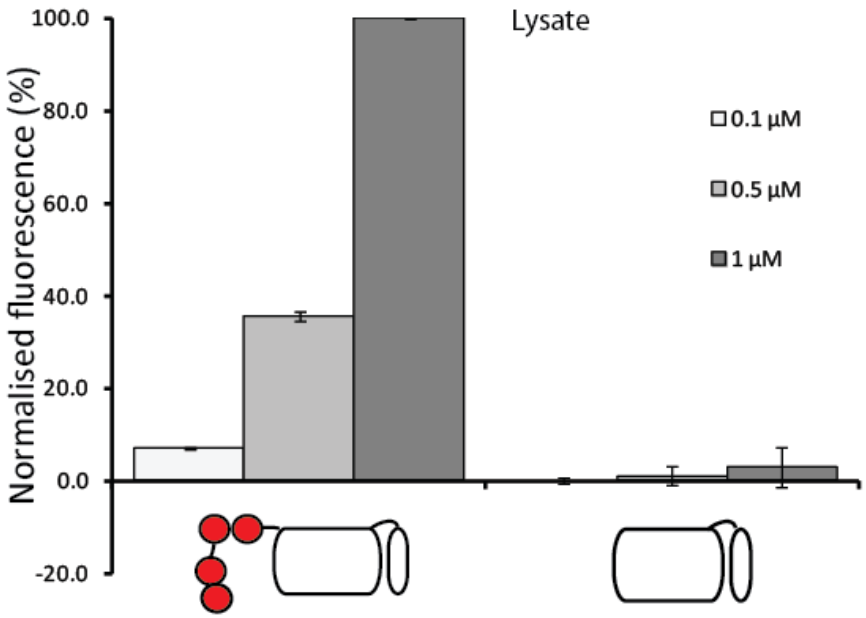


Figure 4.2 – Dose dependent normalised fluorescent signal from both a positive (left - **BfrCCPGCC-N-term**, red dots represent cysteines, bound signal) and negative (right – **BfrWT,** background or free fluorophore signal) control at different FlAsH-EDT_2 concentrations in bacterial lysates resuspended in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8). The only reagent used for this experiment was FlAsH-EDT_2 (1 µM). Total protein concentration: 1 mg/ml. Values are average of six replicates from three different protein expressions. Error bars are S.D.

A number of optimised conditions for working with FlAsH-EDT_2 have been published. These conditions include additives meant to ensure all cysteine residues are fully reduced. Among these are *tris*(2-carboxyethyl)phosphine (TCEP), dithiothreitol (DTT) and 2-mercaptoethanol (2-ME)[20]. 2-ME has also been credited with acting as a shuttling agent to increase the binding kinetics of the probe[15]. In addition the bi-dentate ligand 1, 2-ethanedithiol (EDT) which quenches FLAsH when bound has also been used as an optimisation additive. The buffer additive ethylenediaminetetraacetic acid (EDTA) acts not only to control ion concentration but could also be especially useful in acting as a protease inhibitor in lysate experiments.

To further optimise these conditions for our initial experiments, several combinations of additives were examined to determine their effect on the fluorescence of FlAsH-EDT_2 in the presence of bacterial lysates containing the over expressed positive control, **BfrCCPGCC-N-term** compared to background (Figure 4.3). While a modest difference in fluorescence,

between bound and free fluorophore is observed when only FlAsH-EDT$_2$ is present, this difference is greatly improved with the introduction of additives. The presence of a reducing agent (either TCEP or 2-ME) provided the largest increase in fluorescent signal, suggesting the importance of fully reducing thiols, presumably those associated with cysteines in the binding site. It has been suggested that a high EDT concentration ensures that the arsenic atoms on FlAsH retain bound EDT ligands, and thus quenching the fluorescence, until it is exchanged with two cysteine residues. Therefore it is thought that EDT helps to maintain a low background. In our hands, the addition of EDT caused a small increase in signal but with that followed a small increase in background. The highest fluorescence in the presence of protein with respect to buffer control was observed with addition of all three additives, TCEP, 2-ME and EDT. The combined effect of two reducing agents, one of which could also act as a shuttling agent, and with the bi-dentate ligand EDT, were the optimal conditions to provide the highest positive signal while maintaining a low background.

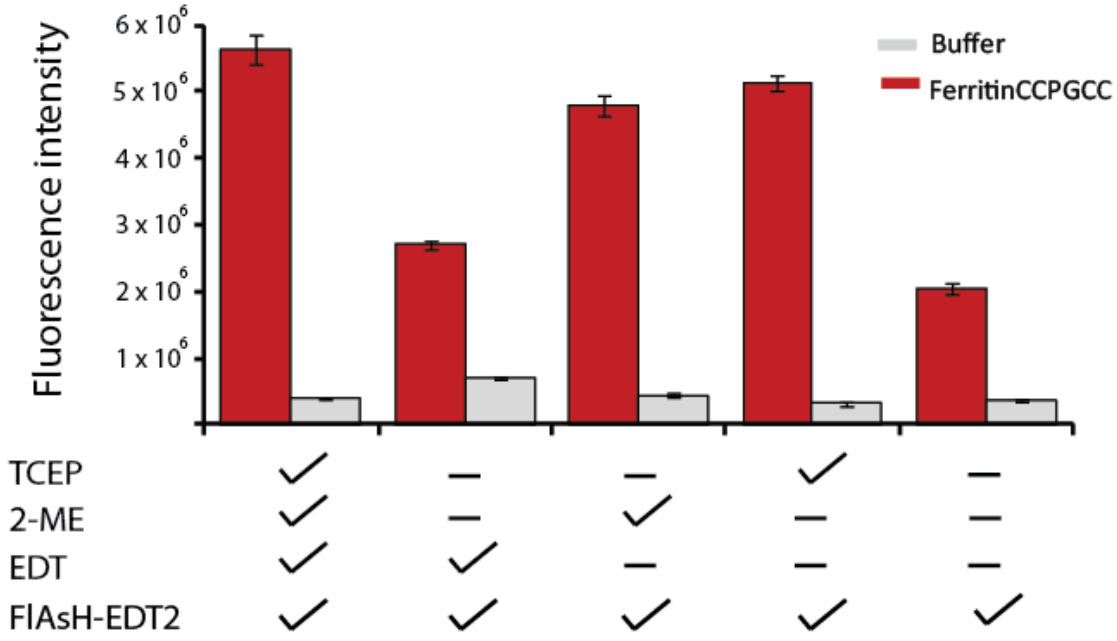


Figure 4.3 – Effect of additives on the fluorescence intensity of FlAsH-EDT$_2$ bound to **BfrCCPGCC-N-term** in lysates with respect to buffer control. Reagent concentrations: TCEP – 3.5 mM, EDT – 1 mM, 2-ME – 1 mM, FlAsH-EDT$_2$ – 1 µM. Total protein concentration: 1 mg/ml in FlAsH buffer, (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) with two hours incubation at room temperature with reagents. Each measurement is an average six replicates from three different protein expressions. Error bars are S.D.

From this experiment, the optimal conditions for a 1 mg/ml protein sample were found to be, 1 µM FlAsH-EDT_2, 3.5 mM TCEP, 1 mM EDT 1 mM 2-ME in buffer containing 100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA at pH 7.8. The additives were added to the protein solution and the solution was incubated for two hours to allow for thorough reduction to occur. The probe was then added and the sample was incubated a further two hours before determining the signal. With controls and conditions set, bipartite designs were designed and analysed.

4.2.2. Design and analysis of bipartite and tetrapartite FlAsH binding sites in Bfr

To achieve the aims of the research described in this chapter, it was necessary to design a binding site for FlAsH-EDT_2 that forms when the ferritin nanocage is assembled, and does not result in the generation of fluorescence upon the formation of other oligomerization states. As stated above, this requirement is especially challenging with Bfr which forms a mixture of dimers and nanocages in solution. Therefore, we need to be sure that we are not only designing binding sites for FlAsH-EDT_2 to detect nanocage formation, but also designing nanocage specific binding sites that do not form upon formation of the dimer.

Another concern that poses a complication for most protein-protein interface engineering studies, but has some distinct aspects for this protein nanocage work, is that an improperly designed binding site might actually disrupt assembly. Therefore we sought a conservative approach in our designs so as to have as little disruption as possible to the secondary, tertiary or quaternary structure. It is thought that the dimer intermediate is similar to the pair of proteins positioned at the two-fold symmetry axis in the nanocage state[4]. This dimer is anti-parallel in that the N- and C-termini of the two proteins diverge. However, in the nanocage, the termini converge at the other symmetry axes. Thus, our strategy focused on placing the partial binding sites at the terminus of the monomers. It was thought that a) upon assembly the termini would converge and generate a cysteine-rich FlAsH-EDT_2 binding site that is specific for the nanocage and not the dimer intermediate b) the binding sites at the termini

would have the least effect on assembly while allowing a degree of flexibility to achieve the optimal projection of cysteines. In the nanocage state, the N-terminus of Bfr points toward a three-fold symmetry axis and is projected outside the nanocage while the C-terminus, is at the end of the E-helix which projects inside of the nanocage at the four-fold axis (Figure 4.4) Therefore, our initial designs focused on placing partial binding sites at these termini.

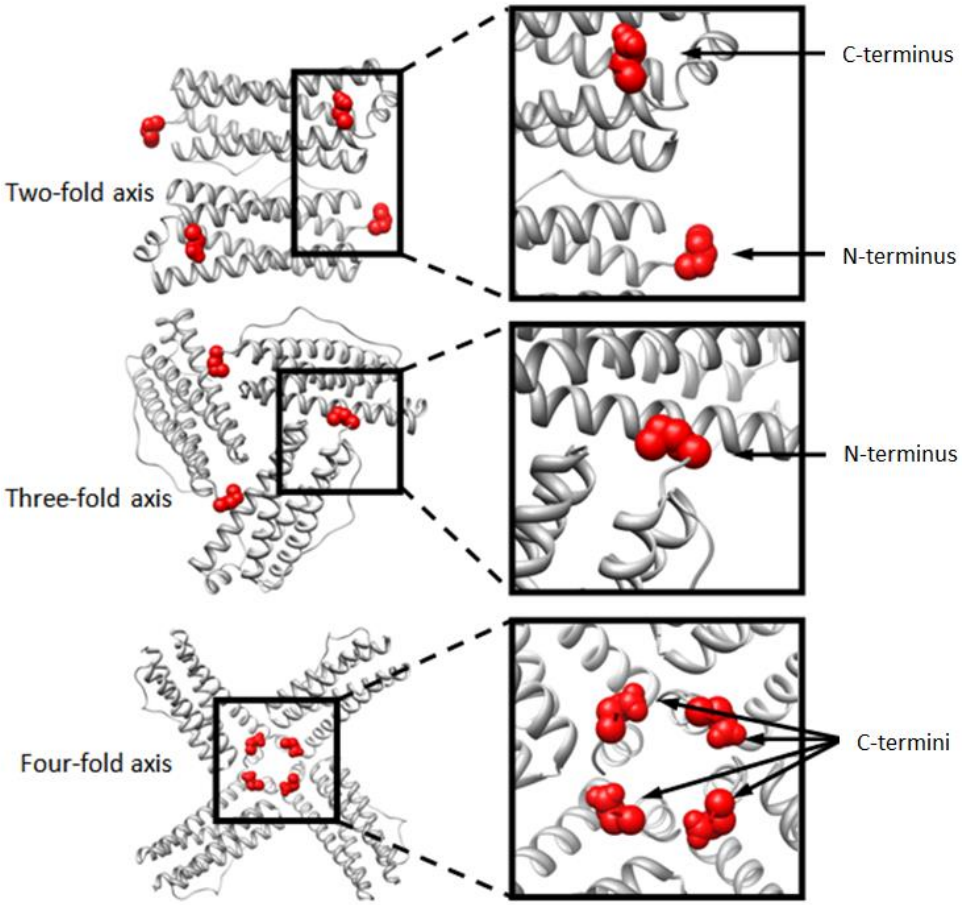


Figure 4.4 – Oligomers of Bfr abstracted from the crystal structure of the nanocage assembly state (PDB:1BFR) emphasizing the two-, three- and four-fold symmetry axes with termini highlighted in red (N-termini at three-fold, C-termini at four-fold and both at two-fold). Images created with Chimera[21].

Multiple designs were generated to increase the likelihood of establishing a successful construct. The first set of designs placed pairs of cysteines at the C-terminus (**BfrCC-C-term**), the N-terminus (**BfrCC-N-term**), and both (**BfrCC-N/C-term**) (Figure 4.5). Because we were curious about the effect of placing a binding site inside the nanocage and because we suspected this nuance might be potentially be consequential for the development of this strategy or might even be the foundation to developing a tool to probe the dynamics of

assembly, an additional positive control was designed by fusing CCPGCC to the C-terminus of Bfr (**BfrCCPGCC-C-term**).

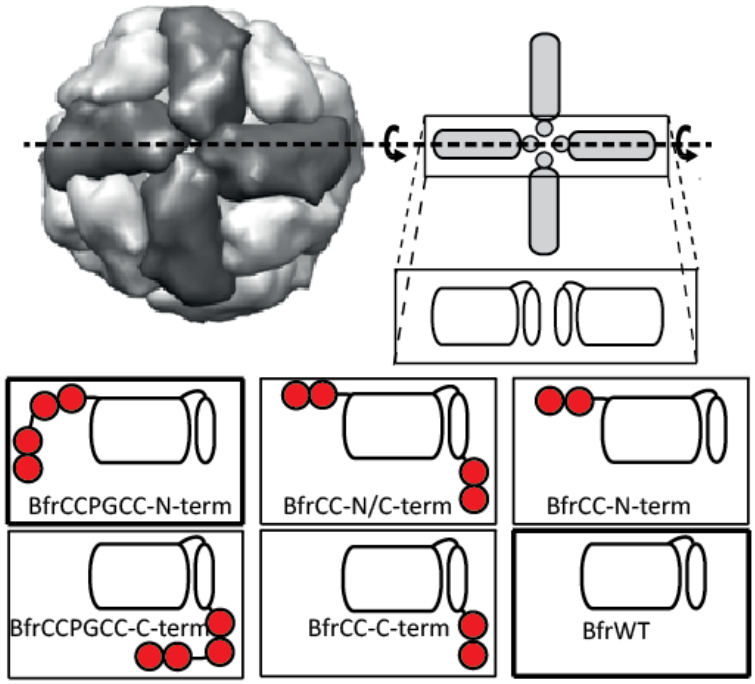


Figure 4.5 – Bipartite Bfr designs for nanocage-dependent FlAsH binding sites. (top left) Bfr crystal structure (PDB: 1BFR) emphasising the four-fold axis of symmetry. (top right) Schematic representing four monomers (elongated) and the E-helix (circle) followed by a side-on view of two of the proteins. (bottom) Engineered monomers viewed side-on with the N-terminus on left and C-terminus on right at the bottom of the E-helix. Red circles indicate cysteine residues. The negative control (**BfrWT**) has no appended cysteines whereas the two positive controls (**BfrCCPGCC-N-term** and **BfrCCPGCC-C-term**) display full binding sites on the monomers.

As an additional design strategy, single cysteine mutations near the Bfr C-terminus were generated to achieve 'tetrapartite' based binding sites upon nanocage formation. As tetrapartite cysteine display was unprecedented, six designs were created along the length of the C-terminal domain containing the E-helix (Figure 4.6). Bfr is ideal for exploration of tetrapartite display because of its four-fold symmetry and that cysteines can be arranged in close proximity at semi rigid locations. Two designs take advantage of the fact that one face of the two turn E-helix faces the other monomers, therefore positions at i (**BfrN148C**) and i+3 (**BfrQ151C**) of the helix were mutated to cysteines. A third design placed a cysteine at the end of the disordered C-terminal tail at the end of the E-helix (**BfrE157C**). Two other designs

extended the C-terminus of Bfr with one cysteine (**BfrExt-C**) and with a GlyCys dipeptide (**BfrExt-GC**). It should be noted that these last two proteins could serve as deletion controls for the **BfrCC-C-term** bipartite design (above). As the C-terminus domain points inside the nanocage, the accessibility to FlAsH-EDT_2 was a concern. A possible solution was to mutate the methionine residue that caps the top of the E-helix to a less bulky alanine in one of the designs, thus generating a double mutant (**BfrM147A,N148C**).

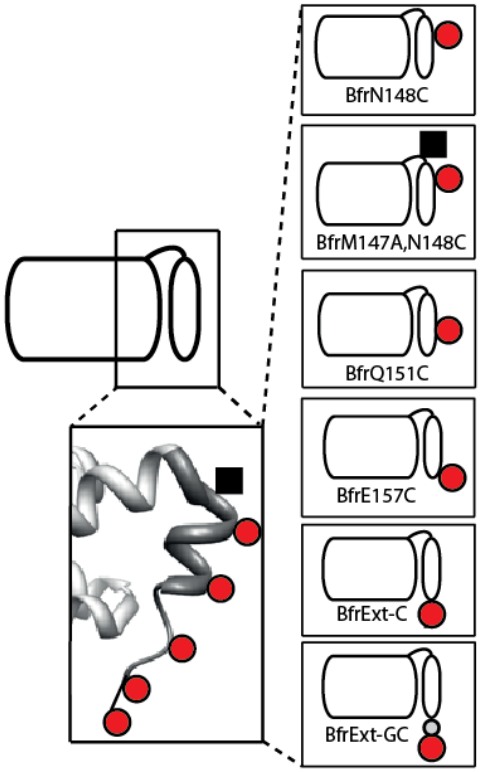


Figure 4.6 – Tetrapartite Bfr designs for nanocage-dependent FlAsH binding sites. (top left) Schematic of Bfr monomer as defined in Figure 4.7. (bottom left) Crystal structure of Bfr (PDB: 1BFR) highlighting the mutated positions. (right) Schematics of designed proteins emphasizing the relative position of mutated residues. (red – cysteine, light grey – glycine, black - alanine).

4.2.2.1. Analysis of bipartite and tetrapartite Bfr designs in lysates

The bipartite Bfr designs (see above), were initially screened in lysates with optimised conditions (see above) (Figure 4.7). As expected and consistent with the optimisation experiments the positive control, **BfrCCPGCC-N-term**, which does not require nanocage formation for the generation of a FlAsH-EDT_2 binding site, displayed a high fluorescence signal upon addition of FlAsH-EDT_2 and the negative control, **BfrWT**, displayed little detectable signal.

These control proteins were used to normalise the fluorescence signal for the other designs. The second positive control, **BfrCCPGCC-C-term**, gave a strong fluorescence although lower than **BfrCCPGCC-N-term**. Interestingly, the N-terminal hexapeptide fusion disrupts nanocage formation whereas **BfrCCPGCC-C-term** forms a substantial amount of nanocage (Figure 4.16). In **BfrCCPGCC-C-term** the binding sites were inside of the nanocage. Taken together, these data suggest that there is a slight inhibition of FlAsH-EDT_2 uptake into the nanocage under these conditions possibility due to the need for the probe to diffuse into the nanocage or due to the need for some nanocage-centered dynamic event (for example, partial disassembly or partial unfolding while assembled etc).

The bipartite designs which showed the highest signals both had cysteine pairs on the C-terminus, with fluorescence around 25% and 20% of the **BfrCCPGCC-N-term** positive control (**BfrCC-N/C-term** and **BfrCC-C-term** respectively) (it should be noted that this response is modest compared to the designs discussed in Chapter 5.). The design containing an N-terminal pair of cysteines, **BfrCC-N-term** resulted in a protein that had no detectable FlAsH-EDT_2 signal. This is not completely unexpected because the crystal structure (Figure 4.4) suggests that the distance between the N-termini of the monomers may be too large (roughly 30 Å) to form a single binding site. As this result suggests that high numbers of additional cysteines in the lysates does not lead to a significant increase in background signal and that the correct geometry is important for FlAsH-EDT_2 fluorescence, this mutant may prove to be a more relevant negative control protein than the wild type. Taken together, these data eliminates the N-terminus of Bfr for bipartite display, but highlights the success of the C-terminal designs.

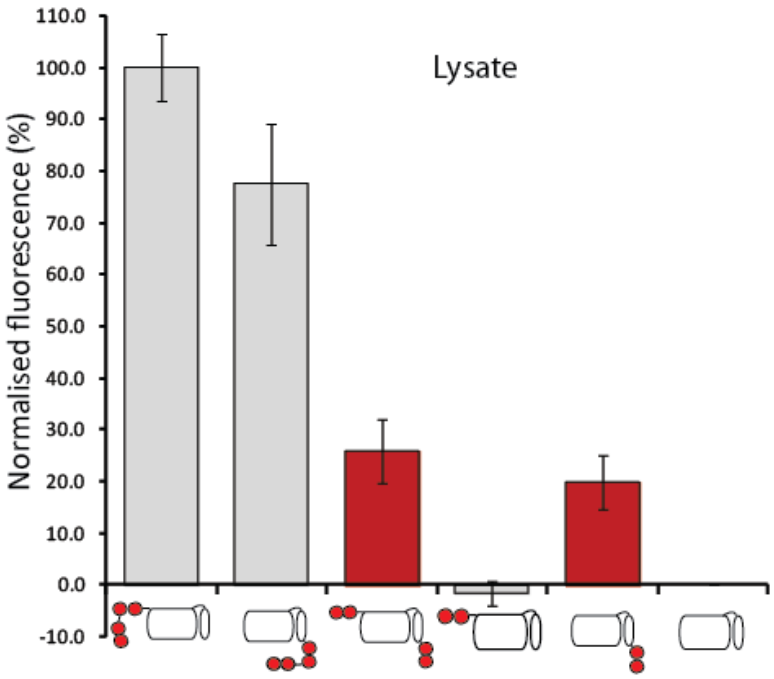


Figure 4.7 - FlAsH fluorescence of Bfr bipartite designs (see figure 4.5 for schematic explanation) in lysates. The data is normalised to the positive and negative controls. All designs set to 1 mg/ml total protein concentration, in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) with 1 µM FlAsH, 3.5 mM TCEP, 1 mM EDT, 1 mM 2-ME, Average of three protein expressions each repeated twice. Error bars are S.D.

The tetrapartite designs were evaluated as described above, using **BfrCCPGCC-N-term** and **BfrWT** as positive and negative controls respectively (Figure 4.8). Interestingly, the two designs, **BfrExt-C** and **BfrExt-GC**, which are the single cysteine versions of **BfrCC-C-term**, showed almost no signal as tetrapartite designs. Negligible signal was observed for **BfrN148C** with little change seen when the methionine was removed, **BfrN147A,N148C**. The highest signals recorded, around 10% of positive control, were from **BfrQ151C** and **BfrE157C.** While this signal is low, it warranted further investigation.

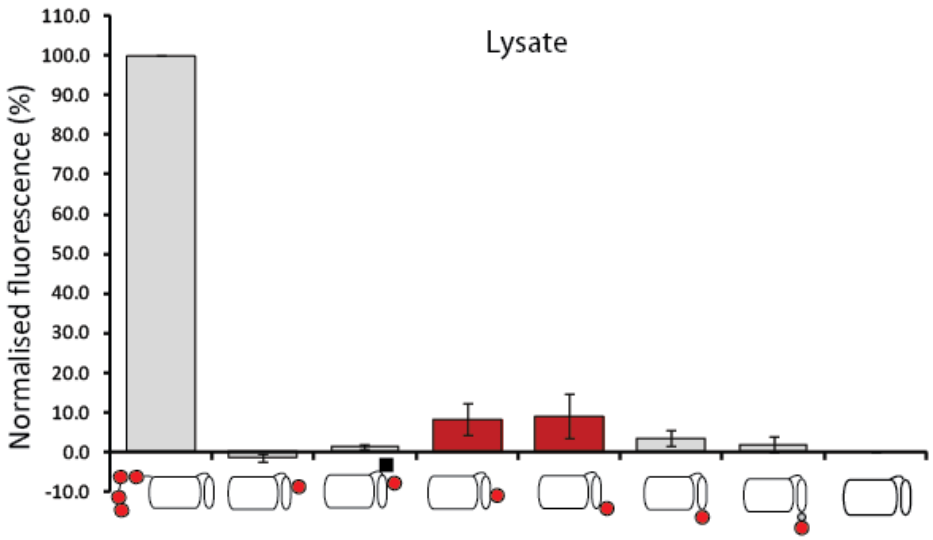


Figure 4.8 - FlAsH fluorescence of Bfr tetrapartite designs (see figure 4.6 for schematic explanation) in lysates. The data is normalised to the positive and negative controls. All designs set to 1 mg/ml total protein concentration, in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) with 1 µM FlAsH, 3.5 mM TCEP, 1 mM EDT, 1 mM 2-ME, Average of three protein expressions each repeated twice. Error bars are S.D.

While this initial data suggests the possibility that several designs are acting as FlAsH-EDT_2 binding sites, it is important to verify that these binding sites are indeed dependent on the self-assembly of the nanocage.

4.2.2.2. Effect of denaturation on Bfr designs in lysate conditions

To determine whether the designed binding sites are self-assembly dependent, the mutants were reanalysed in lysates under denaturing conditions. High concentrations of guanidine·HCl (6 M) were added to each sample in order to remove any quaternary, tertiary and secondary structure (Figure 4.9). Any loss in signal could be attributed to the loss of FlAsH-EDT_2 binding sites. The positive control under denaturing conditions, **BfrCCPGCC-N-term**, demonstrated a high signal, and **BfrWT** had very little signal similar to non-denatured conditions. This is expected because **BfrCCPGCC-N-term** can bind to FlAsH regardless nanocage assembly. Interestingly the other positive control, **BfrCCPGCC-N-term**, showed an observable increase in signal upon denaturation. This result lends further support to the hypothesis that this protein protects the inward pointing site from binding FlAsH-EDT_2 when assembled; upon

denaturation, the site is more accessible. The bipartite designs showing the highest lysate signals were reexamined in denaturing conditions. Both **BfrCC-N/C-term** and **BfrCC-C-term** showed high loss of signal (80 and 90% respectively) as would be expected if the FlAsH binding and fluorescent signal were assembly dependent.

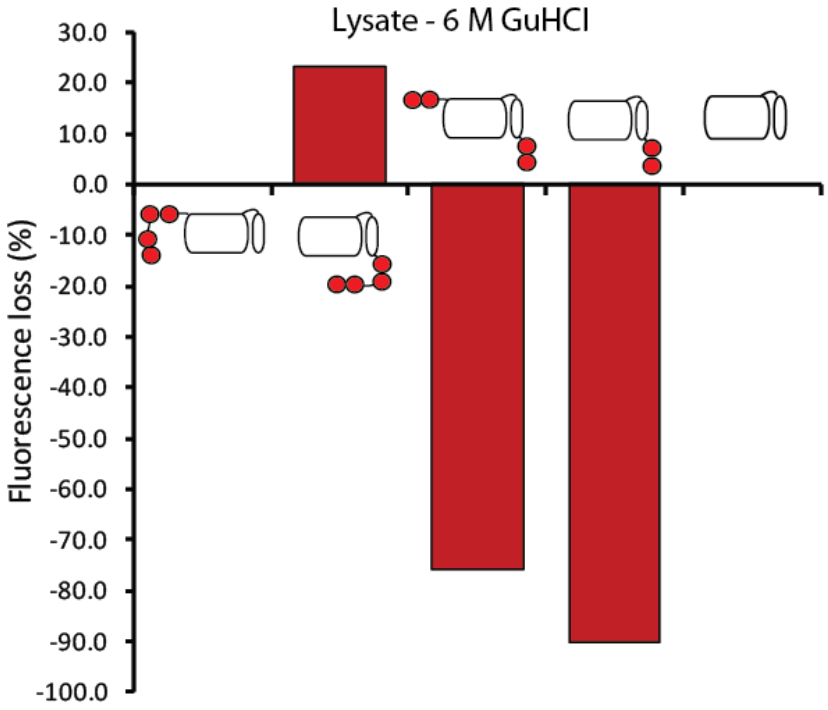


Figure 4.9 – The assembly dependence of FlAsH fluorescence for Bfr bipartite designs (see figure 4.5 for schematic explanation) in lysates under denaturing conditions. The data is normalised to the change in fluorescence of the positive and negative controls. All designs set to 1 mg/ml total protein concentration, in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) with 1 µM FlAsH, 3.5 mM TCEP, 1 mM EDT, 1 mM 2-ME and 6 M guanidine·HCl (GuHCl), Data shows the average change from three protein expressions each repeated twice.

The tetrapartite designs which showed the highest lysate signal were also reanalysed under denaturing conditions (Figure 4.10), and in a similar fashion to the bipartite designs, tetrapartite **BfrQ151C** and **BfrE157C** designs showed significant loss of signal (80 and 100% respectively).

These data taken together suggest that the designs **BfrCC-N/C-term**, **BfrCC-C-term**, **BfrQ151C** and **BfrE157C** provide a binding site in a geometry that can bind FlAsH-EDT_2 and that this binding is oligomerization dependent.

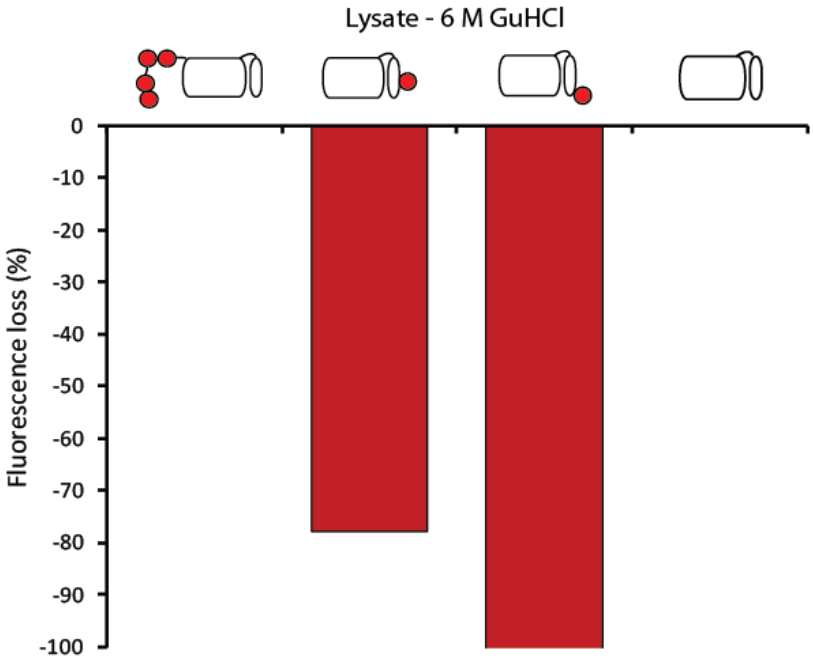


Figure 4.10 - The assembly dependence of FlAsH fluorescence for Bfr tetrapartite (see figure 4.6 for schematic explanation) designs in lysates in denaturing conditions. The data is normalised to the change in fluorescence of the positive and negative controls. All designs set to 1 mg/ml total protein concentration, in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) with 1 µM FlAsH, 3.5 mM TCEP, 1 mM EDT, 1 mM 2-ME and 6 M guanidine·HCl (GuHCl), Data shows the average change from three protein expressions each repeated twice.

4.2.2.3. Detecting stabilising mutations in Bfr

A long term aim of this project is to screen Bfr mutant libraries to discover proteins with enhanced abilities to assemble. Therefore an assay strategy should be robust enough to detect changes in the Bfr oligomerization state ratio. Work conducted by the Orner laboratory to establish rational methods to stabilise Bfr resulted in mutants that favour the nanocage form over the dimer to various degrees (see above). The Bfr mutants, N23F[4] and D118F[10] increased the nanocage populations to roughly 70% and 100% respectively as observed through size exclusion chromatography analysis. Introduction of these mutations into Bfr with designed FlAsH-EDT_2 binding sites would provide a test of whether it is possible to utilise our technique to discover stabilising mutations in the future.

The mutations N23F and D118F were placed onto the best bipartite designs, **BfrCC-N/C-term** (**BfrCC-N/C-termN23F** and **BfrCC-N/C-termD118F**) and **BfrCC-C-term** (**BfrCC-C-termN23F** and

BfrCC-C-termD118F), and analysed in lysates (Figure 4.11). As anticipated, the N23F mutation, which causes wild type Bfr to shift from 50% nanocage to 70%, resulted in an increase in signal in both bipartite backgrounds. The D118F, which pushes wild type entirely to nanocage results in an even more dramatic increase. These data, establish a correlation between the expected concentration of nanocages present in solution and the fluorescent readout paving the way for using this strategy to discover protein cages with enhanced stability. Moreover, these experiments suggest the potential for not only a qualitative assay for the detection of nanocages, but perhaps one that is quantitative.

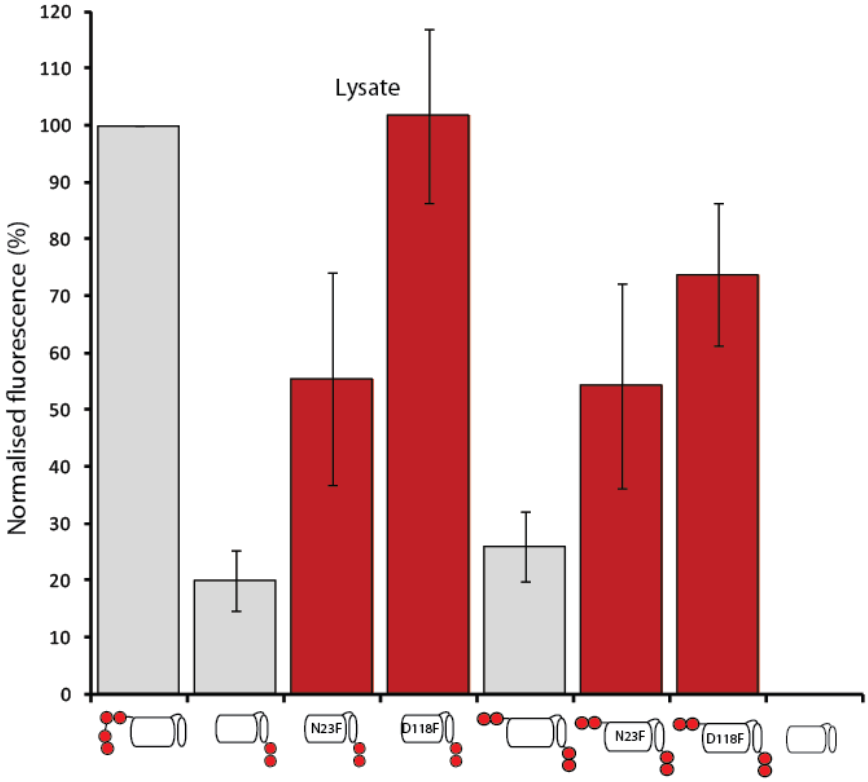


Figure 4.11 – Detecting stabilising mutations with Bfr bipartite designs (see figure 4.5 for schematic explanation) in lysates. The data is normalised to positive and negative controls. All designs set to 1 mg/ml total protein concentration, in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) with 1 µM FlAsH, 3.5 mM TCEP, 1 mM EDT, 1 mM 2-ME. Data from three protein expressions each repeated twice. Error bars are S.D.

The stabilising mutants were also introduced onto the tetrapartite designs, **BfrQ151C** (**BfrQ151C,N23F** and **BfrQ151C,D118F**) and **Bfr157C** (**BfrE157C,N23F** and **BfrE157C,D118F**), and analysed (Figure 4.12). For these designs, the results were less emphatic and somewhat disappointing. While a signal above background was observed for **BfrE157C,N23F**, there was

very little difference compared to its non-stabilised counterpart, and no detectable signal was observed for the other proteins. The dissimilarity between the bipartite and tetrapartite data could be an indication of the unsuitability of tetrapartite designs for rapid screening of Bfr libraries. Their lower lysate signals compared to bipartite designs, also suggest that tetrapartite designs might have a large enough signal to noise ratio window through which to see subtle differences in stability.

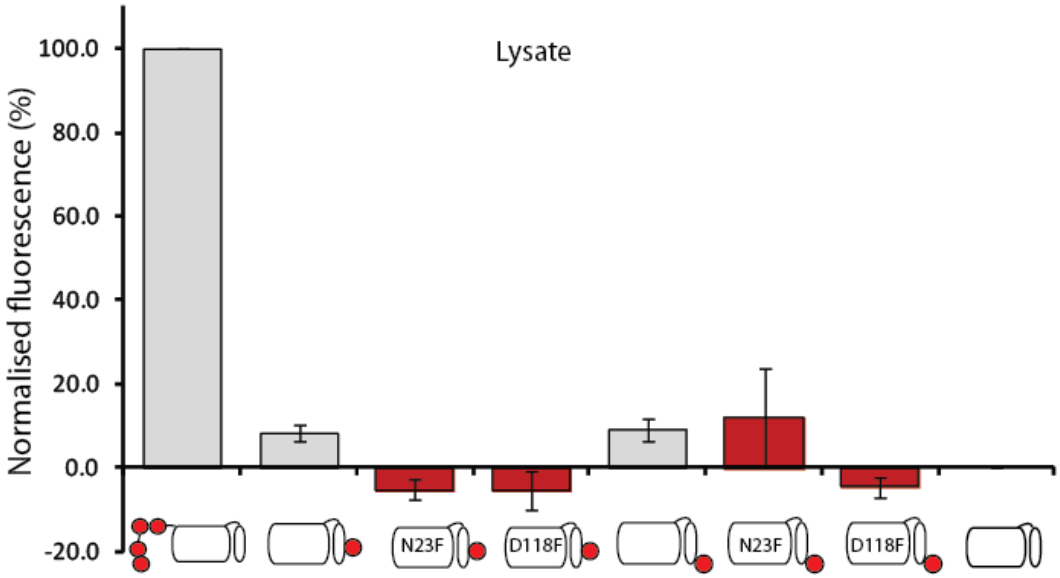


Figure 4.12 - Detecting stabilising mutations with Bfr tetrapartite (see figure 4.6 for schematic explanation) in lysates. The data is normalised to positive and negative controls. All designs set to 1 mg/ml total protein concentration, in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) with 1 µM FlAsH, 3.5 mM TCEP, 1 mM EDT, 1 mM 2-ME. Data from three protein expressions each repeated twice. Error bars are S.D.

The data above suggests that some of the designed systems can a) bind FlAsH b) bind FlAsH in an oligomerization dependent manner and c) be used to distinguish more stable nanocages. However, these experiments were all performed in lysates. While operating in this mode strongly demonstrates its strength in permitting the rapid screening of a number of designs, we need to definitively demonstrate that nanocages are being formed. The lack of

direct observation of the nanocage state limits the definitive conclusions that can be drawn. Therefore we set out to determine the oligomerization state of these proteins.

4.2.2.4. Native polyacrylamide gel electrophoreses (PAGE) analysis of Bfr bi- and tetrapartite designs

It is clear that purified protein samples are required to further interrogate the success of these Bfr designs, as it is essential to determine what oligomerization states are formed by each design as well as providing additional solid proof as to which oligomerization state is responsible for the fluorescent signal. However, expression and purification is a time consuming process and restricts the number of designs that can be assessed. One technique that could bridge the lysate and purified protein work is native PAGE. By running lysate samples of the designs and controls on native gels, the nanocage, dimer and other intermediate species can be separated and observed. Not only would this confirm the presence of nanocages but, if run in the presence of FlAsH-EDT_2, the gel could be visualized by fluorescence to determine the oligomer responsible for the fluorescent signals obtained in lysates.

Lysate samples for the Bfr controls and bipartite designs were analysed via native PAGE with purified **BfrWT** acting as a marker to identify both the nanocage and dimer species (Figure 4.13). Although the resulting data is interesting and can lend insight for the further characterisation of this technique there are some caveats with this data. First, the dimer or possibly some other Bfr intermediate, is clearly running with the gel front as seen when analysing purified Bfr. This, along with data from the dye only lane, which exhibits a strong fluorescence response also at the gel front prevents the ability to clearly identify what species is responsible for fluorescence. This lower sized band, be it a Bfr intermediate or a small molecule is more responsible for the large fluorescence intensity seen in this gel compared to the nanocage. The co-movement of this intermediate and unbound dye prevents precise

identification of the species responsible for this intensity. While contradictory, the data suggests that a gradient native gel would be ideal for the separation of the dimer and the nanocage, the nanocage is readily observable under these conditions. The second caveat is that the nanocage bands exhibit extremely modest fluorescence compared to that arising from the lower co-moving band. Furthermore, in a number of cases, a very large fluorescence band is evident that has no corresponding coomassie band at this point.

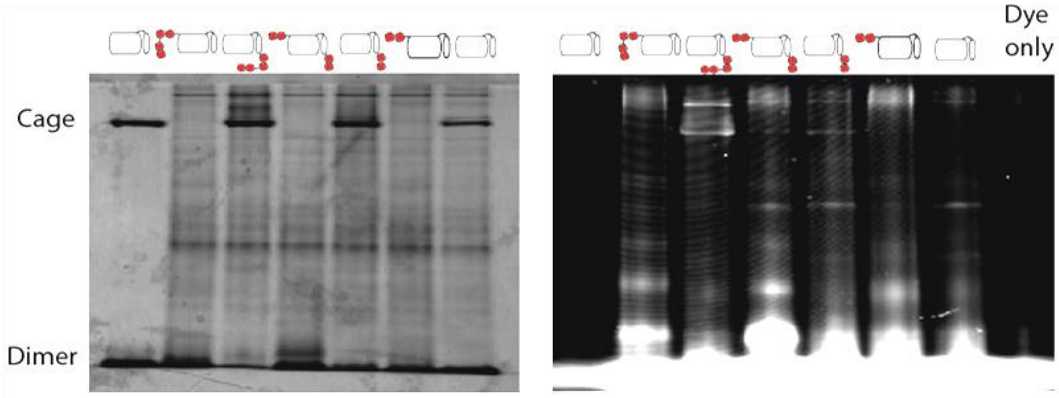


Figure 4.13 – Native PAGE (8%) analysis with both coomassie blue stain (left) and fluorescence (right) visualisation of the same gel, for Bfr bipartite protein designs and controls (see figure 4.5 for schematic explanation). To each well was added 20 µl of overexpressed protein lysate sample set to 1 mg/ml in FLAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) containing 1µM FlAsH-EDT_2, 3.5 mM TCEP, 1 mM EDT and 1 mM 2-ME with purified **BfrWT** being used as a standard (first lane in each gel). Fluorescence visualisation took place using a 488 ± 20 nm excitation filter and a 526 ± 10 nm emission filters.

Another immediate observation from the native gel is that any design with an extension on the N-terminus (**BfrCCPGCC-N-term**, **BfrCC-N-term** and **BfrCC-N/C-term**) showed a complete absence of nanocage bands which suggests that bulky substitution at the N-terminus is detrimental to assembly. Also this electrophoresis data contradicts the initial steady state fluorescence data for **BfrCC-N/C-term** which demonstrated a high signal that was oligomerization dependent (see above).

Designs containing extensions only at the C-terminus (**BfrCCPGCC-C-term**, and **BfrCC-C-term**) showed the presence of nanocages. The presence of nanocages in the **BfrCCPGCC-C-term** sample supports the hypothesis that the reduced signal arising from this protein compared to the positive control (**BfrCCPGCC-N-term**) is due to reduced accessibility to FlAsH-

EDT_2 (above). **BfrCC-C-term** exhibited a strong band corresponding to the presence of nanocage. However the fluorescence imaged gel, shows only a very small corresponding band.

This analysis was repeated for the stabilised mutants **BfrCC-N/C-termN23F**, **BfrCC-N/C-termD118F**, **BfrCC-C-termN23F** and **BfrCC-C-termD118F** in lysate samples (Figure 4.14). For all mutants, a fluorescent, albeit faint, nanocage band was evident. Inconsistent with our previous data (above), however, is that the D118F mutant seems to have less nanocages than the N23F mutant observed through its smaller band size. Under native PAGE scrutiny, the design **BfrCC-N/C-term** exhibited no nanocage band. Results from its stabilised mutants showed the presence of nanocages which corresponded to a faint fluorescent band suggesting the ability of the stabilised mutants to overcome the effects of an extension on the N-terminus.

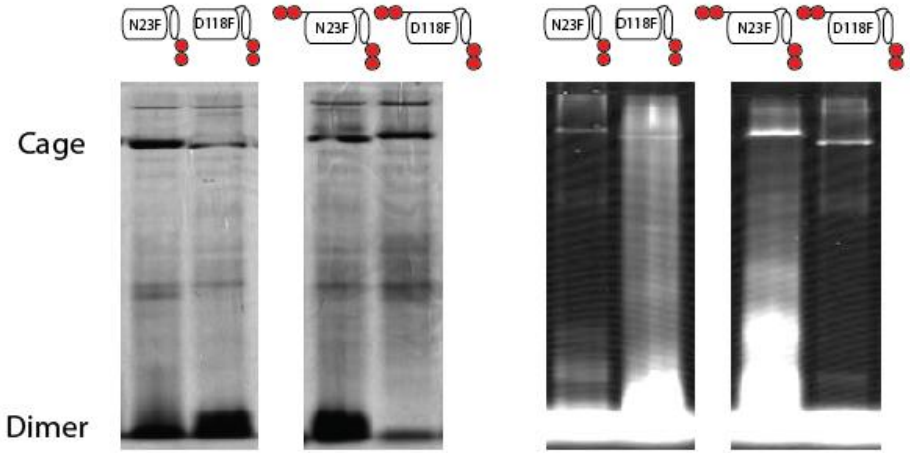


Figure 4.14 – Native PAGE (8%) analysis with both coomassie blue stain (left) and scanned for fluorescence off the same gel (right), for Bfr bipartite protein designs with stabilising mutation N23F and D118F(see figure 4.5 for schematic explanation). To each well was added 20 µl of overexpressed protein lysate sample set to 1 mg/ml in FLAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) containing 1µM FlAsH-EDT_2, 3.5 mM TCEP, 1 mM EDT and 1 mM 2-ME with purified **BfrWT** being used as a standard (first lane in each gel). Fluorescence visualisation took place using a 488 ± 20 nm excitation filter and a 526 ± 10 nm emission filter.

Although the use of native gels was proving problematic, for thoroughness they were used to analyse the Bfr tetrapartite designs (Figure 4.15). Only minor bands corresponding to the nanocage oligomer were observed for **BfrN148C**, **BfrM147A,N148C** and **BfrQ151C** further ruling them out as successful designs, however **BfrE157C**, **BfrExt-C** and **BfrExt-GC** did show the

presence of nanocage bands that could correspond to bands that are fluorescent in the FlAsH gel. Curiously, the coomassie stained gels for **Bfr157C** and **BfrExt-C** exhibited a large band of size intermediate between nanocage and dimer.

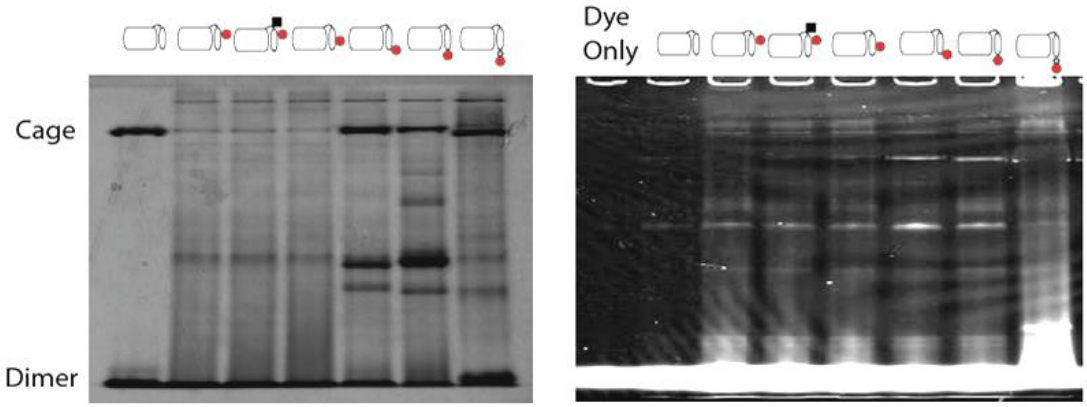


Figure 4.15 – Native PAGE (8%) analysis with both coomassie blue stain (left) and scanned for fluorescence from the same gel (right), for Bfr tetrapartite protein designs and controls (see figure 4.6 for schematic explanation). To each well was added 20 µl of overexpressed protein lysate sample set to 1 mg/ml in FLAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) containing 1µM FlAsH-EDT_2, 3.5 mM TCEP, 1 mM EDT and 1 mM 2-ME with purified **BfrWT** being used as a standard (first lane in each gel). Fluorescence visualisation took place using a 488 ± 20 nm excitation filter and a 526 ± 10 nm emission filter.

While native PAGE, due to the ability to run multiple proteins in parallel, is a relatively rapid method for the observation of Bfr nanocages in lysate samples clearly the technique was producing data that was inconsistent with previous experiments. While it is possible that the issues could be solved through optimisation of conditions, full characterisation of purified proteins was seen as a more precise option. It was also felt that for future understanding of the technology it would be best to understand the pitfalls of this technique it was best to start eliminating the multiple variables associated with lysates.

4.2.2.5. Characterisation of purified Bfr bipartite designs

All controls were expressed and purified along with the two bipartite designs: **BfrCC-N/C-term**, was used to understand why a protein that isn't forming nanocages (see gel data above) can exhibit such a high fluorescence, and **BfrCC-C-term**, which so far, was the best Bfr candidate. The two stabilised mutants, **BfrCC-C-termN23F** and **BfrCC-C-termD118F** which also

could form nanocage oligomers and exhibited high fluorescence intensity in lysates were also expressed and purified. **BfrCC-N-term** and all tetrapartite designs were not generated due to low signal or lack of observable nanocages on the gels.

The proteins described above were purified and fully characterised (See appendix 4.5, 4.6, 4.7, 4.8 and 4.9 for SDS-PAGE, mass spectrometry, CD and TEM and Figure 4.16 for SEC; the experimental details are presented in this chapter below). **BfrCCPGCC-N-term** and **BfrCC-N/C-term** exhibited no nanocage-like oligomers in SEC as expected from the native gel; however, nanocages were observed with TEM. Contradictory results like this are not unknown; we have seen effects with ferritin mutants that form only dimer in solution but can assemble into nanocages in the "forcing", dried conditions of TEM[3]. Consistent with native gels, **BfrCCPGCC-C-term** showed the presence of nanocages in both SEC and TEM as did **BfrWT** and **BfrCC-C-term** along with the stabilised proteins, **BfrCC-C-termN23F** and **BfrCC-C-termD118F**. While **BfrCC-C-termN23F** did not show a significant increase in the nanocage population over background, **BfrCC-C-termD118F** did, shifting the ratio almost completely in favour of nanocages.

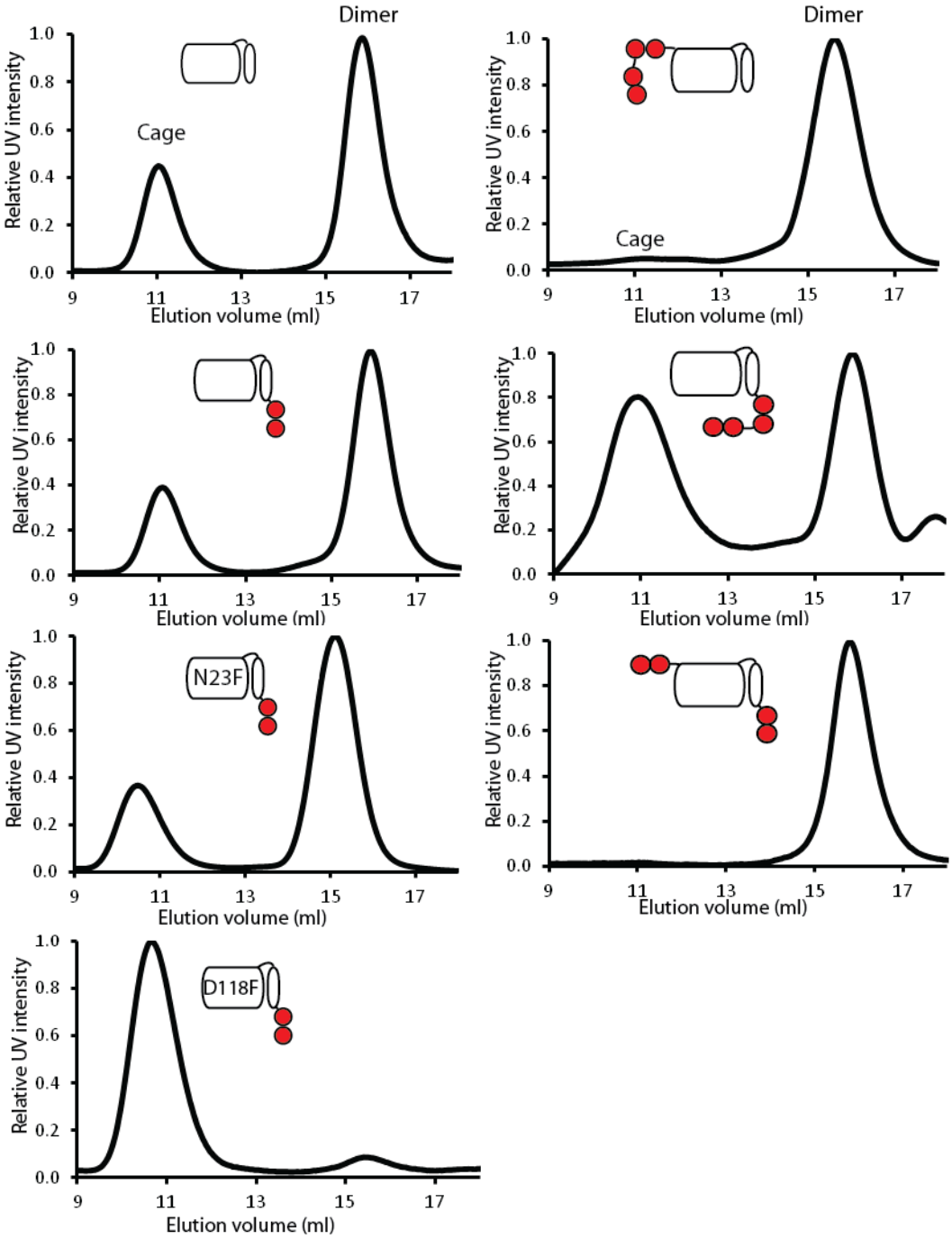


Figure 4.16 - Size exclusion chromatograms of purified Bfr derivatives monitored at absorbance of 280 nm (protein) normalised to their highest intensity. Results for the UV are averages of three runs (0.5 ml injection of 1 mg/ml protein).

With their quaternary structures confirmed, the purified proteins were evaluated for FlAsH binding and fluorescence (Figure 4.17). While the positive and negative controls, as well as **BfrCC-C-term** and **BfrCC-C-termD118F** behaved in a similar manner to lysate experiments (Figure 4.7), confirming that the previously observed fluorescent signals arose from Bfr and not

some other species in the lysate sample, **BfrCC-N/C-term** and **BfrCC-C-termN23F** did not. In this case, **BfrCC-C-termN23F** displayed a decreased fluorescence response compared to the background protein. However, this is consistent with the lack of increase in nanocage population observed in the SEC (see figure 4.16). More surprising, **BfrCC-N/C-term** generated a signal comparable to the positive control, **BfrCCPGCC-N-term**. This is surprising as not only does this design not form any nanocages as evidenced by the SEC, this signal is much higher, compared to controls, to that seen in lysates.

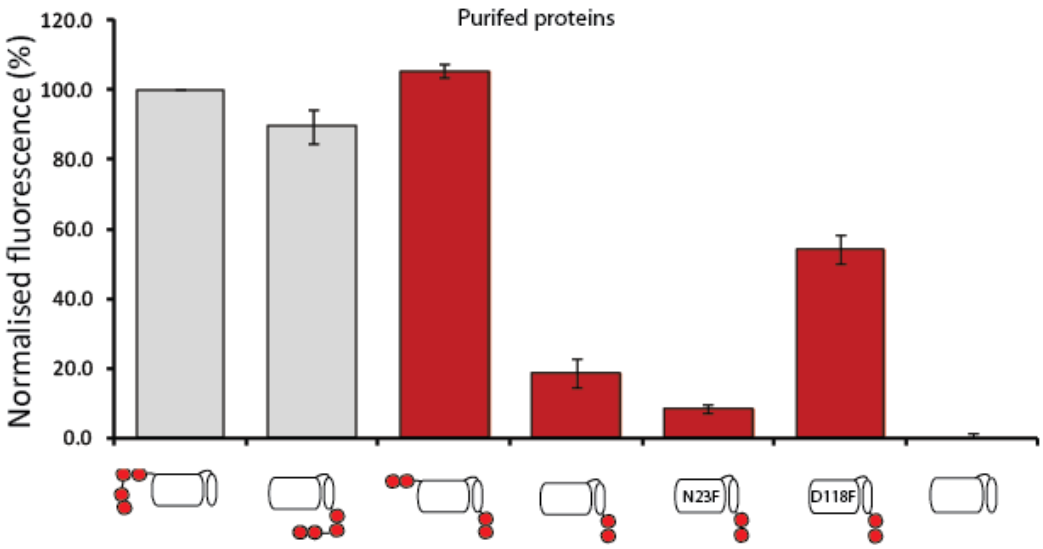


Figure 4.17 - FlAsH fluorescence of purified Bfr bipartite designs (see figure 4.5 for schematic explanation). The data is normalised to the positive and negative controls. All designs set to 1 mg/ml total protein concentration, in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) with 1 µM FlAsH, 3.5 mM TCEP, 1 mM EDT, 1 mM 2-ME. Error bars are S.D. Each purified protein repeated six times.

4.2.2.6. Identification of specific oligomerization states responsible for fluorescence: size exclusion chromatography monitored with a fluorescent plate reader (SEC-FPR)

An aim of the research described in this chapter is to create FlAsH-EDT_2 binding sites in Bfr that can only form when the nanocage assembles. The data obtained so far, especially those of the potential false positive **BfrCC-N/C-term**, highlights that this might be more challenging than first thought. A method to directly detect which oligomer is binding and fluorescing with FlAsH-EDT_2 could be a powerful characterisation tool for future generations of

designs and to verify whether or not **BfrCC-C-term** is acting as a nanocage detector and why **BfrCC-N/C-term** has such a high signal without forming nanocages.

A possible method, in **analogy** to fluorescence-imaged native gels (above), to directly detect which oligomerization state of the protein is responsible for the observed fluorescent signals, could be achieved by combining SEC with the fluorescence assay. By collecting small volume fractions and reading them with a fluorescence plate reader (SEC-FPR), a**n** overlay with the 280 nm absorbance and fluorescence intensity, and thus a correlation between oligomer state and FlAsH signal, could be generated.

Using this technique, the positive control, **BfrCCPGCC-N-term**, exhibited a large fluorescence response co-eluting with the dimer as expected. (Figure 4.18 - a) The negative control, **BfrWT**, had little, although detectable, correlation of fluorescence with the nanocage or dimer peaks (Figure 4.18 – b). It should be noted that both proteins formed a small, rapidly eluting population. It is assumed that this state is nonspecific aggregate. This population has a small amount of detectable fluorescence and, although it is a small population, this could be a concern later on. Another concern is the observation that, in these experimental conditions, the proteins exhibited more oligomerization states than they did in the conditions used for purification (see Fig. 4.16). This effect is mostly likely due to the redox exchange buffer used to optimise FlAsH binding in combination with the disulfide kinetics associated with Bfr assembly. This effect was much less evident with the miniferritin, Dps which was developed in parallel (see Chapter 5). Because of the more direct success of Dps and the desire to expand it into high throughput (see Chapter 6), the conditions for Bfr were never further optimised.

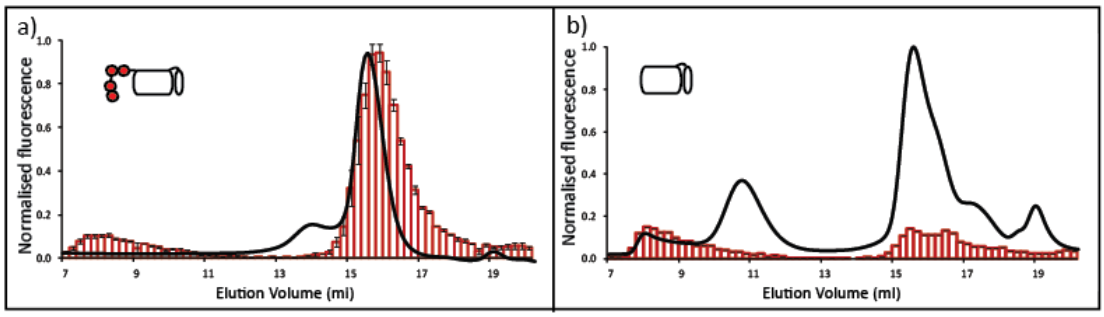


Figure 4.18 – Identification of specific oligomerization states responsible for fluorescence—controls (see figure 4.5 for schematic explanation, a – positive control **BfrCCPGCC-N-term**, b – negative control **BfrWT**). Size exclusion chromatograms monitored at absorbance of 280 nm (protein) and fluorescence samples taken every 200 µl and analysed using a 510 ± 5 nm excitation and a 530 ± 5 nm emission filter set. Each run contained 0.5 ml of 1 mg/ml purified protein in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) with 0.4 µM FlAsH, 3.5 mM TCEP, 1 mM EDT, 1 mM 2-ME. Each protein was repeated three times with data above showing the average of all three runs. Data normalised using the highest gained fluorescent signal as 100% for both sample and the lowest of each as 0%. Error bars are S.D.

The SEC-FPR technique was expanded to the other bipartite designs and controls (Figure 4.19). Most of the proteins showed a strong signal co-eluting with a peak that corresponds to protein aggregates. Worryingly, all C-terminal extensions, **BfrCCPGCC-C-term**, **BfrCC-C-term**, and the corresponding stabilised mutants, showed the presence of nanocage peaks without co-eluting fluorescent signal. Furthermore, strong signals were obtained from peaks corresponding to the dimer. Taken together, these data suggest that proteins with poorly designed or ill-placed FlAsH-EDT_2 binding sites or proteins with oligomerization kinetics on the order of FlAsH exchange could result in spurious responses from nonspecific aggregation, artifactual oligomerization or other false negative/positive generating mechanisms. Many valuable lessons were learned from working with this system however the miniferritin, Dps came progressed with more straightforward data (see Chapters 5 and 6). Although initial experiments and controls appeared positive for Bfr, further characterisation proved intractable. The application of bipartite tetracysteine display for a nanocage oligomerization assay to Bfr was therefore placed on hold.

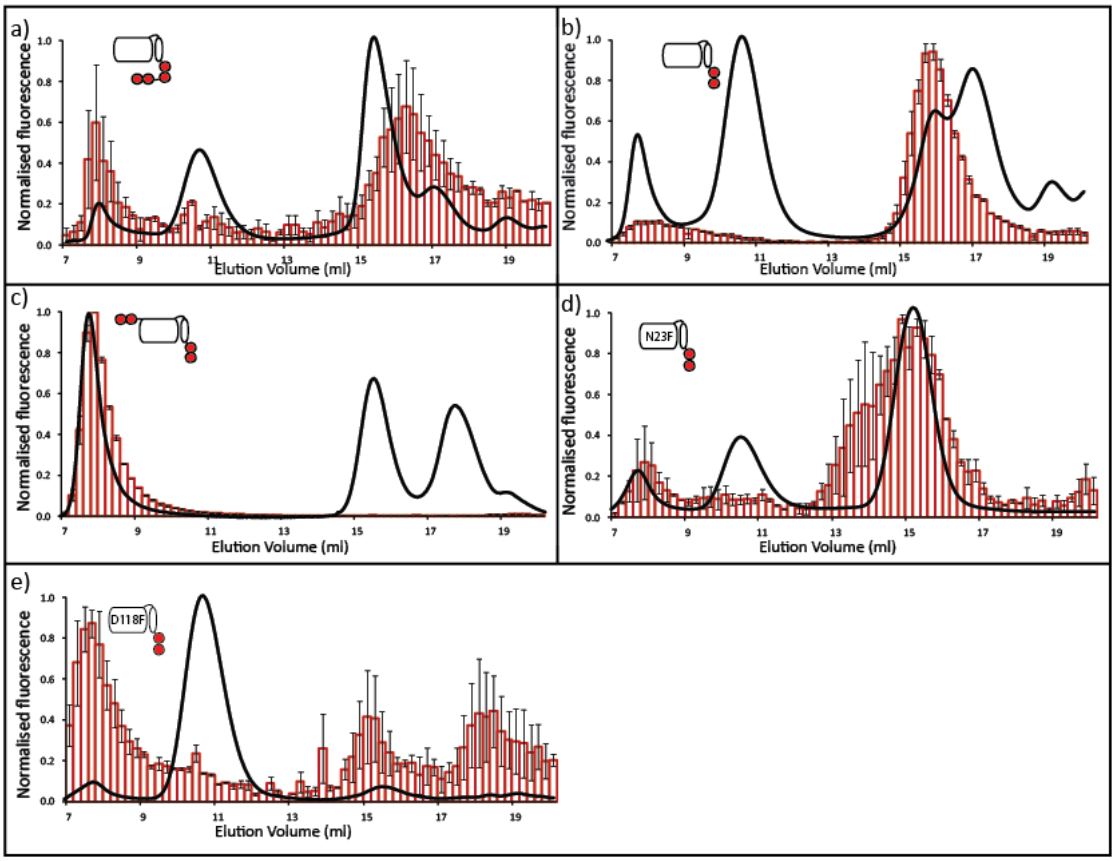


Figure 4.19 - Identification of specific oligomerization states responsible for fluorescence—designs (see figure 4.5 for schematic explanation, a – **BfrCCPGCC-C-term**, b – **BfrCC-C-term,** c – **BfrCC-N/C-term,** d – **BfrCC-C-termN23F** and e – **BfrCC-C-termD118F**). Size exclusion chromatograms monitored at absorbance of 280 nm (protein) and fluorescence samples taken every 200 µl and analysed using a 510 ± 5 nm excitation and a 530 ± 5 nm emission filter set. Each run contained 0.5 ml of 1 mg/ml purified protein in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) with 0.4 µM FlAsH, 3.5 mM TCEP, 1 mM EDT, 1 mM 2-ME. Each protein was repeated three times with data above showing the average of all three runs. Each design was normalised individually using the highest and lowest fluorescent signals as 100% and 0% respectively. Error bars are S.D.

4.2.3. A potential new direction: Detection of Bfr assembly through binding site inaccessibility

Although some of the reasons for the failure of the Bfr designs remain unclear, the data above, while emphasizing the importance of rigorous control experiments, establishing useful characterisation methods, and beginning to define the scope of the technology, lends some additional fundamental insight. The results taken together suggest that the efficiency of the C-terminus to bind FlAsH-EDT_2 is restricted due to its position inside the protein nanocage. Although this fact could, in some regard, be a limitation, it was realized that it could also be exploited to assess nanocage formation in a reversed mode

BfrCCPGCC-C-term exhibits a decreased fluorescence signal under both lysate (Figure 4.7) and purified conditions (Figure 4.17) compared to **BfrCCPGCC-N-term**. Denaturing **BfrCCPGCC-C-term** (Figure 4.9) recovers the signal. It is thought that this effect is due to slower FlAsH-EDT_2 penetration kinetics. Providing a complete binding site that is inaccessible to FlAsH-EDT_2 only in the assembled, nanocage state could be the basis for a method to detect nanocage formation and thus decreased signal would correlate to increased assembly. Because all proteins would have a complete, four cysteine binding motif, problems associated with bipartite and tetrapartite design could be bypassed.

To test this concept, the stabilising mutants N23F and D118F were placed into the control BfrCCPGCC-C-term, which has a complete binding site for each monomer (**BfrCCPGCC-C-termN23F** and **BfrCCPGCC-C-termD118F**) (Figure 4.20). Because **BfrCCPGCC-N-term** exists in solution as a dimer (Figure 4.16) and is therefore freely accessible to FlAsH-EDT_2, this protein could act as a positive signal/negative assembly control. **BfrCCPGCC-C-term**, would serve as the negative signal/positive assembly control. A more stabilising mutant would generate less signal than both of these controls but have a great percentage increase in signal upon denaturation.

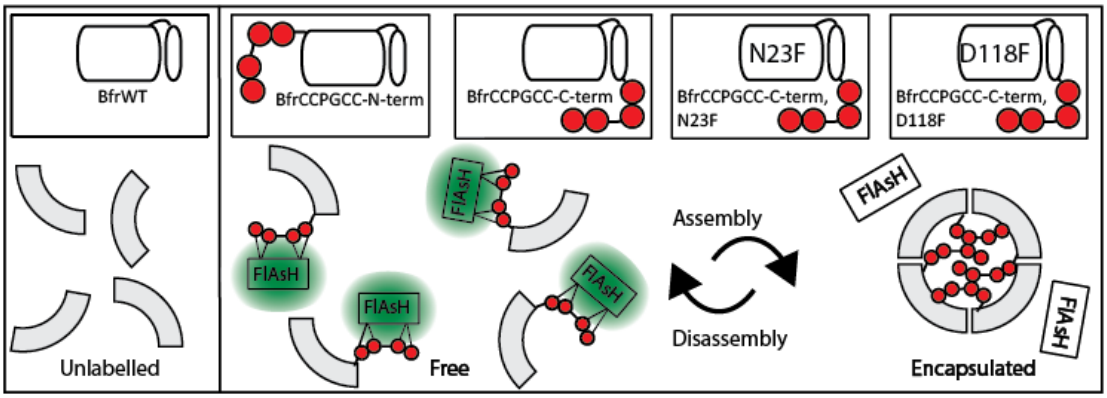


Figure 4.20 – Detection of Bfr assembly through binding site inaccessibility. (Bottom) Schematic demonstrating how the change in accessibility of full, four cysteine binding sites can be decreased upon nanocage assembly if the binding sites are positioned inside the cage. (Top) designs of controls and stabilised nanocages to test this concept (red dots – cysteine amino acids).

It was realized that the signal response window is narrow and the number of binding sites is increased, so for this system to be successful, further optimisation of the protein and probe stoichiometry is essential. In addition, so as to be more accurate in controlling the number of binding sites, we only worked with purified proteins. To that end, the proteins **BfrCCPGCC-N-term**, **BfrCCPGCC-C-term**, **BfrCCPGCC-C-termN23F**, **BfrCCPGCC-C-termD118F** and **BfrWT** were expressed, purified and characterised (see Appendices 4.5, 4.6, 4.7, 4.8, 4.9 and 4.10 for mass spectrometry, SDS-PAGE, TEM and SEC and section 4.4 of this chapter for experimental procedures).

As expected, **BfrCCPGCC-C-termN23F** and **BfrCCPGCC-C-termD118F** showed nanocage character in both SEC and TEM with an increase in nanocage population compared to **BfrCCPGCC-C-term**. These proteins were assayed for their fluorescence intensity with a 1:1 or a 2:1 probe to protein molar ratio (Figure 4.21). For both ratios, the trend in signal was as expected from the least stable nanocage, **BfrCCPGCC-C-term**, generating the highest signal and the other proteins having decreased signal correlated to their relative stabilities with the 1:1 ratio providing the largest range in fluorescence (Fig. 4.21a and 4.21c) However, less expected were the results from the denaturation control where it was thought that upon denaturation, all the binding sites would be freely accessible to the probe and the signals would match that of **BfrCCPGCC-N-term.** This was clearly not the case where the stabilised mutations exhibit nearly undetectable changes in fluorescence upon denaturation. (Fig. 4.21 b and 4.21d) In fact, the control protein **BfrCCPGCC-N-term,** did not reach full signal upon denaturation at the 2:1 probe to protein ratio suggesting that something was clearly wrong with these conditions. Further conditions were analysed using different concentrations of proteins and fluorophore, with similar trends observed (see appendix 4.11). At the moment it remains unclear what is going on with this experiment and further optimisation is required. As these experiments were undertaken, the Dps designs were working well and were pursued while Bfr was placed on hold.

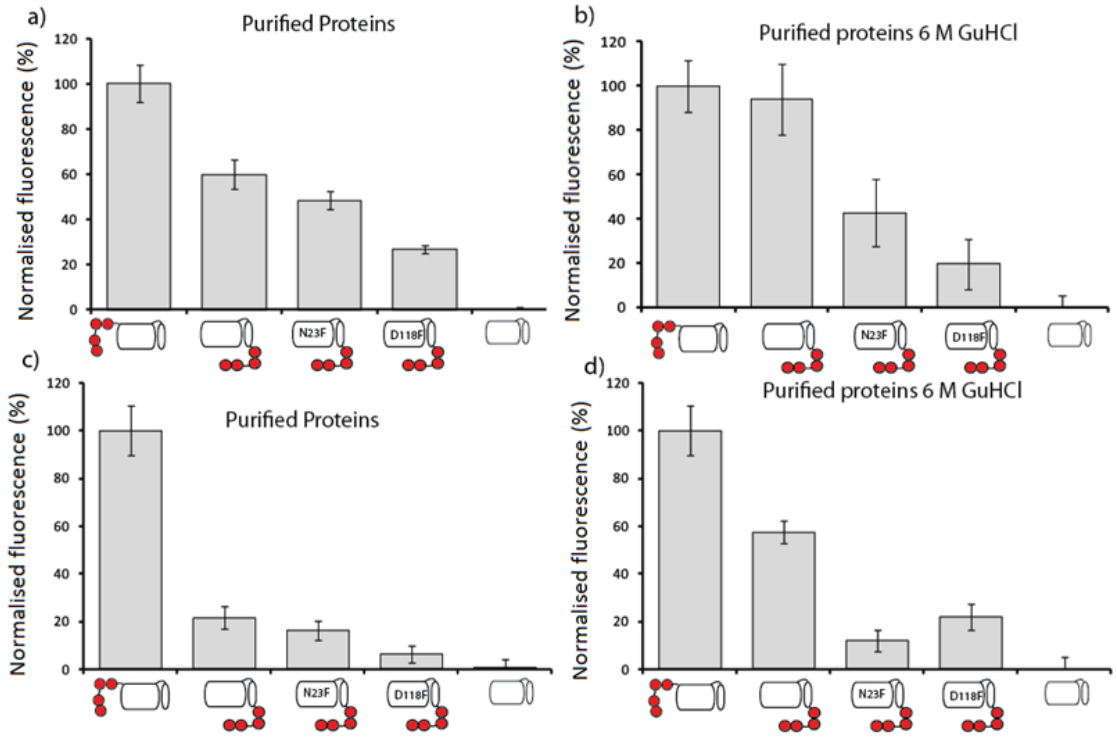


Figure 4.21 – Fluorescence analysis to detect Bfr assembly through binding site inaccessibility. (a and c) normalized fluorescence of purified proteins (b and d) recovery of fluorescence through denaturation. (a and b) 1:1 molar ratio of fluorophore to monomer (c and d) 2:1 ratio of fluorophore to monomer. The data is normalised to positive and negative controls. All designs set to 0.1 mg/ml total protein concentration, in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) with, 3.5 mM TCEP, 1 mM EDT, 1 mM 2-ME and varying concentrations of FlAsH-EDT_2. Denatured experiments contained 6 M GuHCl. Each experiment repeated 6 times and error bars are S.D.

4.3. Conclusion

A method to rapidly and directly determine the self-assembly of bacterioferritin using bipartite tetracysteine display and encapsulated full binding sites was developed and analysed. All designs were constructed around the rule that any binding site could bind the probe FlAsH and fluoresce only upon nanocage assembly formation and not with any other oligomerization state. Selective fluorescence signals from both bipartite and tetrapartite designs were observed in lysate conditions suggesting the flexibility and robustness of these designs. These signals could be switched off upon protein denaturation suggesting that they are oligomerization dependent. Furthermore, by applying mutations that have been previously shown to stabilise nanocage formation, the FlAsH signal was greatly enhanced. Further analysis, using native PAGE and SEC with purified proteins was performed to confirm the oligomerization state responsible for observed fluorescent signals. A new technique was developed that combined size exclusion chromatography with a fluorescence plate assay. However, when examined, none of the designed proteins exhibited any fluorescent signal associated with the protein nanocage and showed multiple oligomerization states. Fluorescent signals from the designs instead resulted from either dimer intermediates or protein aggregates. As an alternative strategy, a second method relying on the shielding of the full FlAsH binding site inside the assembled nanocage was developed. While purified protein fluorescence signals showed trends similar to those expected, there was little change observed upon denaturation suggesting a lack of correlation between oligomerization state and fluorescent signal.

It is unknown what caused the failure of the bipartite and tetrapartite designs. One possibility could be related to the fact that many more peaks in the SEC under FlAsH binding conditions than when the protein was initially purified. This suggests that complex, disulfide exchange to generate transient but non-native oligomerization may be occurring. Clearly more optimisation of conditions needs to be performed. What is also unknown is why the bipartite

designs showed dimers that appeared to fluoresce in these conditions. It is possible that either the oligomerization state is misassigned in the SEC or that these dimers are not native dimers because it is unclear how our designs would permit fluorescence when the dimers are antiparallel, two-fold symmetric with divergent termini observed in the crystal structure.

As we learned that the buffer conditions may be important to the success of this technique, it is thought that the kinetics of both binding and assembly should also be a focus of future optimisation. The majority of Bfr designs on the C-terminus placed the binding sites inside the nanocage. For the incubation times used in our standard conditions, positioning the binding sites inside the nanocage resulted in restricted access by FlAsH. However, while attempting to exploit this setback into a new strategy of probing assembly by purposely restricting access to full binding sites, problems where encountered in our denaturation controls. To work as designed, all binding sites should be exposed and able to bind FlAsH-EDT_2 resulting in an increased signal upon denaturation. However, little change in fluorescence was observed. Incomplete denaturation could be the cause although this is unlikely due to the high concentrations of denaturant (6 M GuHCl) and extended denaturation time (6 hours). A more probable explanation again may lie with non-native oligomerization further emphasizing the importance of buffer and protein/probe stoichiometry optimisation.

From these experiments it was learned that placing binding sites at exposed termini can generate FlAsH binding sites that are oligomerization dependent and that this technique can be robust in complex lysates. The characterisation methods and designed rules established here can be easily ported to other designs and the robustness in lysates points the way toward work in living cells which will be the focus of the following two chapters.

4.4. Methods and Materials

4.4.1. Cloning of Bfr designs and controls

4.4.1.1. Transfer of Bfr gene to pET-22b plasmid for lysate analysis with extensions to N-terminus for BfrCCPGCC-N-term, BfrCC-N-term and BfrWT designs

The **BfrWT** gene was amplified with primers providing restriction sites and the required alterations on the N-terminus[2]. The PCR reaction (50 µl total volume) was a solution of reaction buffer (Fermentas, 5 µl of 10xPfu buffer), dNTP mix (Fermentas, 2 µl of a solution containing dATP, dTTP, dGTP, and dCTP at 2 mM each), forward and reverse primers (400 ng each, Suprenom, Singapore, see Appendix 4.1), template (100 ng) and Pfu polymerase (Fermentas, 2 µl of 2.5 U/µl) which was subjected to an initial melting step at 95 °C for 30 s, followed by 30 cycles of amplification (95 °C for 30 s, 55 °C for 45 s and 72 °C for 1 min) followed by 72 °C for 7 min. The resulting PCR product was isolated by gel purification (Promega). The PCR product (300 ng) was digested (Tango buffer, Fermentas, 10 µl of 2x buffer), Nde1 (Fermentas, 1.5 µl of 20 U/µl), Xho1 (Fermentas, 1.5 µl of 20 U/µl), 50 µl total volume, 37 °C, 3 hours). The plasmid pET-22b (Novagen) was treated in the same way, followed by ligation of the insert with T4 ligase buffer (Fermentas, 2 µl, 10x buffer), digested insert (30 ng), digested plasmid (80 ng), T4 ligase (Fermentas, 0.2 µL of 20 U/µl), 20 µl total volume (37 °C, 2 hours). The constructs (2 µl) were transformed (Novablue, Novagen) and the resulting colonies were assessed by colony PCR (a solution of buffer (2.5 µl of 10x), $MgCl_2$ (Fermentas, 1.75 µl of 25 mM) dNTP mix (Fermentas, 2 µl of a solution containing dATP, dTTP, dGTP, and dCTP at 2 mM each), T7 promoter and terminator primers (150 ng each, Suprenom, Singapore), Taq DNA polymerase (Fermentas, 0.25 µl of 5 U/µl), colony suspension (10 µl of a 50 µl total colony suspension in deionised water) and deionised water (5 µl), was subjected to an initial melting step (95 °C for 10 min) followed by 30 cycles of amplification, ((95 °C for 30 s, 59 °C for 30 s and 72 °C for 30 s) followed by 72 °C for 7 min). Clones demonstrating a PCR product of the predicted size were

subjected to miniprep (Invitrogen) and the resulting purified plasmids were sequenced (see Sequence Appendix 4 for sequence results and Appendix 4.1 for primers used).

4.4.1.2. Construction of BfrCCGPCC-C-term, BfrCC-C-term, BfrCC-N/C-term, BfrExt-C, BfrExt-GC, BfrCC-C-termN23F and BfrCC-C-termD118F via site directed mutagenesis for screening in lysates

The constructs for **BfrCCPGCC-N-term**, **BfrCC-N-term** and **BfrWT** in pET-22b were used as templates for site directed mutagenesis (Quikchange, Stratagene) to introduce mutations at the C-terminus. The reaction mixture containing Pfu turbo buffer (Fermentas, 2 µl of 10x), dNTP mix (Fermentas, 2 µl of a solution containing dATP, dTTP, dGTP, and dCTP at 2 mM each), forward and reverse primers (125 ng each, Suprenom, Singapore, see appendix 4.1), the dsDNA template (100 ng) and Pfu turbo polymerase (Fermentas, 1 µl of 2.5 U/µl) in a total volume of 25 µl was subjected to melting (95 °C for 5 min), followed by 20 cycles of amplification (95 °C for 1 min, 55 °C for 1 min and 68 °C for 6 min) followed by 70 °C for 10 min. The reaction was then digested with Dpn1 (NEB buffer 4 (1 µl of 10x), amplification product (5 µl), deionized water (3.5 µl), Dpn1 (NEB, 0.6 µl of 10 U/µl), 37 °C, 2h). The resulting solution was desalted (Promega, PCR cleanup kit) and transformed. The resulting colonies subjected to miniprep (Invitrogen) and sequenced (see Sequence Appendix 4 for sequence results and Appendix 4.1 for primers used).

4.4.1.3. Expression plasmid transfer of Bfr designs to allow for analysis with purified proteins

In order to express the proteins for isolation and purification, the genes were placed into a vector that provides affinity tags. The constructs were amplified from the mutated pET-22b plasmid with primers providing ligation independent cloning (LIC) sites. The PCR solution (Pfu reaction buffer (Fermentas, 5 µl of 10x), dNTP mix (Fermentas, 2 µl of a solution containing dATP, dTTP, dGTP, and dCTP at 2 mM each), forward and reverse primers (400 ng each, Suprenom, Singapore, see Appendix 4.1), reverse oligonucleotide primer Suprenom, Singapore), the template (100 ng) and Pfu polymerase (Fermentas, 2 µl of 2.5 U/µl) in 50 µl

total volume) was subjected to an initial melting step (95 °C for 30 s), followed by 30 cycles of amplification (95 °C for 30 s, 55 °C for 45 s and 72 °C for 1 min) followed by 72 °C for 7 min. The resulting PCR product was isolated by gel purification (Promega). The PCR product was then treated with T4 polymerase in order to create the required complementary overhangs (NEB buffer 2, 2 µl of 10x buffer), dATP (NEB, 2 µl of 25 mM), DTT (Sigma, 1 µl of 100 mM), BSA (NEB, 0.2 µl of 100x), PCR product (0.3 pmol) and T4 DNA polymerase (NEB, 0.6 µl of 10 U/µl) in a total volume of 20 µl 30 min, 22 °C). The insert was annealed to open plasmid pET-32b (Novagen) (1 µl of the T4 Pol product, 0.5µl of plasmid solution (50 ng/µl)) by incubating (room temperature, 30 min) followed by a second incubation (room temperature, 30 min) with EDTA (1 µl of 100 mM). The constructs (2.5 µl) were transformed (Novablue, Novagen) and the resulting colonies were assessed by colony PCR (a solution of buffer (2.5 µl of 10x), $MgCl_2$ (Fermentas, 1.75 µl of 25 mM) dNTP mix (Fermentas, 2 µl of a solution containing dATP, dTTP, dGTP, and dCTP at 2 mM each), T7 promoter and terminator primers (150 ng each, Suprenom, Singapore), Taq DNA polymerase (Fermentas, 0.25 µl of 5 U/µl), colony suspension (10 µl of a 50 µl total colony suspension in deionised water) and deionised water (5 µl), was subjected to an initial melting step (95 °C for 10 min) followed by amplification 30 cycles of (95 °C for 30 s, 59 °C for 30 s and 72 °C for 30 s) followed by 72 °C for 7 min). Clones demonstrating a PCR product of the predicted size were subjected to miniprep (Invitrogen) and the resulting purified plasmid was sequenced. For **BfrCC-C-term**, **BfrCC-N/C-term** and the stabilized mutants N23F and D118F, site directed mutagenesis was performed as described above using specific primers for the required mutations (see Sequence Appendix 4 for sequencing results and Appendix 4.1 for primers used).

4.4.1.4. Construction of Bfr designs with restricted full binding sites, BfrCCPGCC-N-term, BfrCCPGCC-C-term, BfrCCPGCC-C-termN23F and BfrCCPGCC-C-termD118F

Site directed mutagenesis was performed on the **BfrCCPGCC-C-term** design in pET-32b with the primers necessary for N32F and D118F mutations as described in 4.4.1.2. (see sequence Appendix 4 for sequence results and Appendix 4.1 for primers used)[4, 10].

4.4.2. FlAsH analysis in lysates

The pET-22b expression vectors containing Bfr variants were transformed into BL21 *E.coli* cells and plated on LB plates (50 µl/ml of carbenicillin). Selected colonies then were grown in LB (3 ml with 50 µl/ml of carbenicillin, 37 °C, overnight) as a pre-culture which was then added to LB (100 ml) and grown (37 °C) until an $O.D_{600}$ of 0.6. Protein expression was then induced by the addition of IPTG (400 mM) and the culture was further incubated (3 h, 30 °C with the exception of **CC-C-termN23F** and **CC-C-termD118F** (20 °C, overnight). The cells were isolated by centrifugation (10,000 rpm, 20 min at 4 °C). The cell pellet was resuspended in Flash buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) and sonicated (Sonics vibra cell sonicator, 60% amplitude, pulsed 5 s on 5 s off for 15 min). The protein solution was clarified by centrifugation (10,000 rpm, 20 min, 4 °C) and then filtered (Pall, 0.2 µm). The protein concentration was determined (BCA, Novagen) and the cell lysate was diluted to 1 mg/ml. with FlAsH buffer. To assure that all the 1 mg/ml samples had similar amounts of the desired protein, each was analyzed by SDS PAGE (see appendix 4.2, 4.3 and 4.4). For each sample, (200 µl, 1 mg/ml) in FlAsH buffer TCEP (Sigma, final concentration of 3.5 mM), EDT (Sigma, final concentration of 1 mM) and 2-ME (Sigma, final concentration of 1 mM) were added and the solution was incubated (room temperature, 2 h) followed by the addition of FlAsH-EDT_2 (Invitrogen, final concentration of 1 µM) followed by a further incubation (20 °C, 2 h) in the dark. Each lysate sample was examined (Varian Cary Eclipse spectrofluorimeter, Ex 510 nm, Em 530 nm, Ex and Em slits 5 nm, for 1 min in kinetic mode in a quartz fluorescence cuvette) at least six times in total with at least three separate protein expressions assessed twice. For the denaturing experiments each sample was also incubated with a final concentration of 6 M guanidine at the same time as the addition of TCEP, EDT and 2-ME.

4.4.3. FlAsH analysis with purified proteins

The pET-32b vectors containing the Bfr variants were transformed into BL21 *E.coli* cells and plated on LB plates (50 µl/ml of carbenicillin). Selected colonies were then grown in LB (5 ml, 37 °C, overnight) as a pre-culture which was added to LB (500 ml) and grown (37 °C) until an $O.D_{600}$ of 0.6. Protein expression was then induced by the addition of IPTG (final concentration of 400 mM) and the cultures were further incubated (3 h, 30 °C). The cells were isolated by centrifugation (10,000 rpm, 20 min, 4 °C). The cell pellet was resuspended in lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 40 mM Imidazole, pH 8) and sonicated (Sonics vibra cell sonicator, 60% amplitude pulsed 5 s on, 5 s off for 15 min). The protein solution was clarified by centrifugation (10,000 rpm, 20 min at 4 °C in an Eppendorf 5810 centrifuge) and then filtered (Pall, 0.2 µm). The protein was purified via affinity (GE, Histrap FF, 5 ml, (wash buffer-40 mM Imidazole, 50 mM NaH_2PO_4, 300 mM NaCl, pH 7.4), (elution buffer-500 mM Imidazole, 50 mM NaH_2PO_4, 300 mM NaCl, pH 7.4)). Enteriokinase digestion (NEB 2 µg/ml) was performed to cleave off the peptide tag from the protein of interest followed by a second Histrap (GE, Histrap FF, 5 ml, (wash buffer-40 mM Imidazole, 50 mM NaH_2PO_4, 300 mM NaCl, pH 7.4), (elution buffer-500 mM Imidazole, 50 mM NaH_2PO_4, 300 mM NaCl, pH 7.4)) to remove the tag from solution. The protein solution was further purified by size exclusion chromatography (GE Hiload 16/60 Superdex). The degree of purification was assessed with SDS PAGE (see appendix 4.5), SEC (see figure 4.18), TEM, CD and mass spec (appendix 4.6, 4.7, 4.8, 4.9 and 4.10). FlAsH analysis was performed as described above for lysate proteins, with each sample being repeated 8 times with a set concentration of 1 mg/ml. Bfr encapsulated designs were analysed with different concentrations of both fluorophore and protein as indicated in the figure captions above.

4.4.4. Protein characterisation

4.4.4.1. Size exclusion chromatography (SEC) and SEC-FPR

The samples (0.5 mg/ml) in GFC buffer (50 mM Na_2HPO_4, 100 mM NaCl, pH 7) was injected (0.5 ml) on to the column (GE Superdex 200 10/300 GL) at 0.5 ml/min and each was repeated 3 times. The column was calibrated using six proteins as standards (GE Biosystems Calibration Kit).

For fluorescent monitoring, samples were prepared by adding to pure protein (0.5 ml of 1 mg/ml) in FlAsH buffer, TCEP (Sigma, final concentration of 3.5 mM), EDT (Sigma, final concentration of 1 mM) and 2-ME (Sigma, final concentration of 1 mM) and incubated (2 h, room temperature). FlAsH-EDT_2 dye was added and incubated in the dark (Invitrogen, final concentration of 0.4 µM, 2 h, room temperature). This sample was desalted (GE, HiTrap 5 ml Desalting column) before injection onto a SEC column (GE Superdex 200 10/300 GL). Samples were taken every 200 µl during the elution and placed into a black Corning 96 well plate. This plate was examined in a PerkinElmer Envision 2101 multilabel plate reader (Ex filter 485 nm bandwidth 14 nm, Em filter 535 nm bandwidth 25 nm). Each purified protein was analysed three times with each plate re-read three times.

4.4.4.2. Native gel fluorescence analysis

Native gels were performed (4% stacking gel, 8% resolving gel, 20 min 80 volts, followed by 1 h 130 volts at 4 °C) on all samples subjected to FlAsH analysis (total protein concentration of 1 mg/ml in 200 µl total volume with 1 µM FlAsH-EDT_2). Each gel was scanned with a fluorescent gel reader (Typhoon Trio Variable mode imager Ex 488 nm, Em 526 nm) and then stained with coomassie brilliant blue. Gel pictures were analyzed using ImageJ[23].

4.4.4.3. Transmission Electron Microscopy (TEM)

TEM was performed on a FEI, Tecnai G^2 20, electron microscope set at 200 KeV. Purified proteins (0.01 mg/ml in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8)

were immobilized on Formvar/carbon coated 3.05 mm copper grids (TAAB) and negatively stained with 1% Uranyl acetate. Micrographs were analyzed using ImageJ[23] (see Appendix 4.9).

4.4.4.4. Circular Dichroism Spectroscopy (CD)

All proteins were purified and made to 0.2 mg/ml in phosphate buffer (50 mM Na_2HPO_4, 50 mM NaCl, pH 7) performed on a JASCO J-810 spectropolarimeter from 250 nm to 200 nm with a path length of 0.5 mm. Thermal melts were performed from 5 to 100 °C [24, 25] (see Appendix 4.7).

4.4.4.5. Mass Spectrometry

Purified proteins (0.05 mg/ml) were desalted into dH_2O (GE, Hitrap 5ml Desalting column) and analysed on a Bruker MaXis mass spectrometer after the addition of 1% Formic acid (see Appendix 4.6).

4.5. References

1. Cornell, T. A.; Orner, B. P., Detection of Protein Cage Assembly through Probing with BisArsenic Fluorophores. *Meth Mol Biol* 2014, *Accepted*.
2. Fan, R.; Boyle, A. L.; Cheong, V. V.; Ng, S. L.; Orner, B. P., A Helix Swapping Study of Two Protein Cages. *Biochemistry* 2009, *48* (24), 5623-5630.
3. Zhang, Y.; Raudah, S.; Teo, H.; Teo, G. W. S.; Fan, R.; Sun, X.; Orner, B. P., Alanine-shaving Mutagenesis to Determine Key Interfacial Residues Governing the Assembly of a Nano-cage Maxi-ferritin. *Journal of Biological Chemistry* 2010, *285* (16), 12078-12086.
4. Ardejani, M. S.; Li, N. X.; Orner, B. P., Stabilization of a Protein Nanocage through the Plugging of a Protein-Protein Interfacial Water Pocket. *Biochemistry* 2011, *50* (19), 4029-4037.
5. Zhang, Y.; Fu, J.; Chee, S. Y.; Ang, E. X. W.; Orner, B. P., Rational disruption of the oligomerization of the mini-ferritin E. coli DPS through protein-protein interface mutation. *Protein Science* 2011, *20* (11), 1907-1917.
6. Aisen, P.; Listowsky, I., Iron transport and storage proteins. *Annual Review of Biochemistry* 1980, *49*, 357-393.
7. Theil, E. C., Ferritin - structure, gene-regulation, and cellular function in animals, plants, and microorganisms. *Annual Review of Biochemistry* 1987, *56*, 289-315.
8. Harrison, P. M.; Arosio, P., Ferritins: Molecular properties, iron storage function and cellular regulation. *Biochimica Et Biophysica Acta-Bioenergetics* 1996, *1275* (3), 161-203.
9. Dautant, A.; Meyer, J. B.; Yariv, J.; Precigoux, G.; Sweet, R. M.; Kalb, A. J.; Frolow, F., Structure of a monoclinic crystal form of cytochrome b1 (bacterioferritin) from E-coli. *Acta Crystallographica Section D-Biological Crystallography* 1998, *54*, 16-24.
10. Ardejani, M. S.; Chok, X. L.; Foo, C. J.; Orner, B. P., Complete shift of ferritin oligomerization toward nanocage assembly via engineered protein-protein interactions. *Chemical Communications* 2013, *49* (34), 3528-3530.
11. Cunningham, B. C.; Wells, J. A., High-resolution epitope mapping of hgh-receptor interactions by alanine-scanning mutagenesis. *Science* 1989, *244* (4908), 1081-1085.
12. Zhang, Y.; Orner, B. P., Self-Assembly in the Ferritin Nano-Cage Protein Superfamily. *International Journal of Molecular Sciences* 2011, *12* (8), 5406-5421.
13. Khare, G.; Nangpal, P.; Tyagi, A. K., Unique Residues at the 3-Fold and 4-Fold Axis of Mycobacterial Ferritin Are Involved in Oligomer Switching. *Biochemistry* 2013, *52* (10), 1694-1704.
14. Fields, S.; Song, O.-k., A novel genetic system to detect protein–protein interactions. *Nature* 1989, *340* (6230), 245-246.
15. Griffin, B. A.; Adams, S. R.; Tsien, R. Y., Specific covalent labeling of recombinant protein molecules inside live cells. *Science* 1998, *281* (5374), 269-272.
16. Adams, S. R.; Campbell, R. E.; Gross, L. A.; Martin, B. R.; Walkup, G. K.; Yao, Y.; Llopis, J.; Tsien, R. Y., New biarsenical Ligands and tetracysteine motifs for protein labeling in vitro and in vivo: Synthesis and biological applications. *Journal of the American Chemical Society* 2002, *124* (21), 6063-6076.
17. Griffin, B. A.; Adams, S. R.; Jones, J.; Tsien, R. Y., Fluorescent labeling of recombinant proteins in living cells with FlAsH. *Applications of Chimeric Genes and Hybrid Proteins Pt B* 2000, *327*, 565-578.
18. Ignatova, Z.; Gierasch, L. M., Monitoring protein stability and aggregation in vivo by real-time fluorescent labeling. *Proceedings of the National Academy of Sciences of the United States of America* 2004, *101* (2), 523-528.

19. Luedtke, N. W.; Dexter, R. J.; Fried, D. B.; Schepartz, A., Surveying polypeptide and protein domain conformation and association with FlAsH and ReAsH. *Nature Chemical Biology* 2007, *3* (12), 779-784.
20. Goodman, J. L.; Fried, D. B.; Schepartz, A., Bipartite Tetracysteine Display Requires Site Flexibility for ReAsH Coordination. *Chembiochem* 2009, *10* (10), 1644-1647.
21. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* 2004, *25* (13), 1605-1612.
22. Fan, R.; Boyle, A. L.; Vee, V. C.; See, L. N.; Orner, B. P., A helix swapping study of two protein cages. *Biochemistry* 2009, *48* (24), 5623-5630.
23. Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W., NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* 2012, *9* (7), 671-675.
24. Kelly, S. M.; Jess, T. J.; Price, N. C., How to study proteins by circular dichroism. *Biochimica Et Biophysica Acta-Proteins and Proteomics* 2005, *1751* (2), 119-139.
25. Greenfield, N. J., Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols* 2006, *1* (6), 2876-2890.

Chapter 5

Direct detection of mini-ferritin assembly with the fluorescent probe FlAsH-EDT_2

NOTE: Parts of this chapter were taken from published work with the candidate as first author (see appendix – Published Work 2)[1].

5.1. Introduction

The complexity of creating a specific protein nanocage self-assembly detection strategy using FlAsH-EDT_2 technology, was emphasized by the initial attempts with the maxi-ferritin Bfr (see Chapter 4). Although fluorescence signals were obtained for several bipartite and tetrapartite designs and these were shown to be oligomerization dependent, the data did not hold up to rigorous scrutiny, and, in the conditions explored, we observed signal associated with protein aggregates and dimers. However, control proteins and tools were established (for example SEC-FPR) that could complement prolonged biophysical characterisation. It was therefore thought, that it was worth determining whether a similar strategy would be more successful with a different protein nanocage.

5.1.1. Aims of this investigation

With lessons learned from Chapter 4, an altered approach was taken using a different protein nanocage. This nanocage needed to be less complex than Bfr but still be a multimeric, highly symmetric protein nanocage. A protein nanocage that natively assembles into a single species, should be easier to work with and be less prone to the complexities seen with Bfr. We had also learned that placing the FlAsH binding sites on the outside of the nanocage might avoid additional complications. The goals of the research described in this chapter remain the same as in Chapter 4: Develop the ability to analyse multiple conditions or mutants of a protein nanocage for the stability of its quaternary structure stability with the long term aim of transferring this technique from *in vitro* to *in vivo* where protein nanocage libraries could be screened using cell sorting. Perhaps by using a target protein with more tractable properties, we can more directly establish basic and fundamental guidelines to help develop a nanocage-focused FlAsH-EDT_2 technology.

5.1.2. New model system for a protein self-assembly assay: Favourable aspects of Dps

As discussed in Chapter 2, the *E.coli* DNA-binding protein from starved cells (Dps), is a mini-ferritin with a 9 nm outer diameter that assembles from twelve identical monomers and stores cellular iron through mineralisation inside its hollow cavity. Each monomer folds into a four helix bundle with an additional helix in the loop between the second and third helix of the bundle (the "BC helix")[2, 3]. Dps self-assembles into a nanocage with tetrahedral symmetry, possessing six identical two-fold symmetry axes, most likely the dimer intermediate, and four each of two distinct three-fold axes. During Dps self-assembly, monomers rarely persist and as for most ferritins, a two-fold symmetric dimer is believed to be the most prevalent intermediate[4]. This dimer is thought to be the fundamental building block for nanocage formation, however, unlike Bfr, the dimer species in Dps represents a nearly undetectable fraction of the total protein population when analysed with SEC[3]. However, like Bfr, these dimer intermediates are presumed to be anti-parallel, meaning that the termini of one monomer projects away from those of the other monomer; only with increased oligomerization to the fully formed nanocage state do the termini converge.

Similar to Chapter 4, the aim of the project described in this chapter is to create a method to detect nanocage formation with specificity over all other assembly intermediates. Dps was selected for this work, as this mini-ferritin, not unlike Bfr, assembles around symmetric protein-protein interfaces, some of which are nanocage- and not dimer-specific. The low level of dimer in native Dps also might make is more tractable model system than Bfr. In addition both termini of Dps are located on the outside compared to the complicating inward projection of the Bfr C-terminus (see Chapter 4). Therefore it was thought that Dps could be a superior model system to Bfr (Figure 5.1).

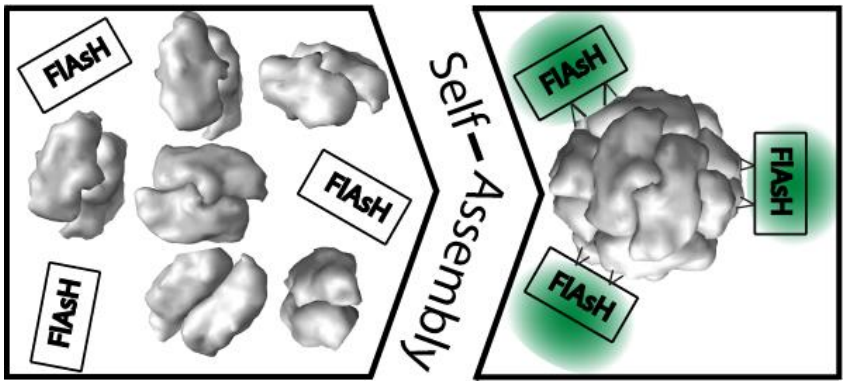


Figure 5.1 – A self-assembly assay for Dps protein nanocage self-assembly should generate signal in a nanocage specific manner. In an ideal assay, the fluorescent probe, FlAsH should bind and fluoresce, only in the presence of the nanocage oligomerization state and not the anti-parallel dimer. Designed binding sites for FlAsH should thus only form upon nanocage formation.

5.1.3. FlAsH binding site design strategies applied to Dps

One strategy to design binding sites in Dps is to focus on the termini in analogy to what was done in Chapter 4 with Bfr. A potential advantage to Dps is that both the N- and C-terminal ends are located outside of the nanocage. The most accessible convergence for FlAsH binding site design is formed by the C-terminus which is located at a three-fold axis of symmetry. The N-terminus comprises of a long unstructured flexible loop (residues 1-21) with only the last few amino acids observable in most crystal structures (PDB:1DPS, 1L8H and 1JRE)[2]. The unstructured nature of this loop could prove useful, as mutations located here would not interfere with secondary structure, lowering the risk of disrupting nanocage assembly. However, it is unclear where the end of this terminus is located and whether or not it converges with other termini upon nanocage formation. Thus the N-terminus was deemed less ideal than the C-terminus for binding site design.

Along with designs focused on the convergence of Dps termini, other strategies were employed. The designs using the termini as set out in Chapter 4 relied on the idea that concentrating cysteines in space would lead to a FlAsH binding site. No real concern was given to stoichiometry and geometry. Therefore, more precise designs were conceived. Schepartz and co-workers have established geometric design rules for optimal FlAsH-EDT_2 binding[5, 6] with separation distances between cysteine sulfur atoms guiding binding site construction. This

geometrically optimised binding site approach was used to aid in unravelling the mechanism of Epidermal growth factor receptor (EGFR) transmembrane signalling[7]. We therefore had this work in mind when designing second generation FlAsH binding sites in Dps.

5.2. Results and discussions

5.2.1. Dps designs

Two philosophies were pursued in designing FlAsH binding sites for Dps. In one, the interaction between one dimer subunit and other dimers in the assembled nanocage acted as a starting point. In the other, the goal was to generate binding sites by converging cysteines in space upon nanocage assembly through N- and/or C-terminal fusions. The fundamentals of this latter philosophy was discussed and employed in Chapter 4. To pursue the former philosophy, all dimer-dimer interactions that form in the Dps nanocage were surveyed for their potential to be exploited for FlAsH binding sites. High potential interfaces were determined by three rules: 1) Mutatable residues at the interfaces must be close in space (<8 Å), 2) Mutatable residues cannot be in between the same dimer unit, 3) Mutatable residues must be predominantly on the outside of the nanocage. These rules, formed from lessons learned in Chapter 4 and derived from the desire to have an accessible binding site that would not form independently of the nanocage oligomerization state. The size constraints on the distance between amino acids followed the principles set out by Schepartz and co-workers by designing binding sites based on the geometry of FlAsH itself[5]. These rules were kept intentionally flexible and do not exclude the termini-focused designs. Because we had the luxury of being able to assess the designs in lysates, we could explore a wider range of designs and scrutinize these two philosophies in parallel. It should be mentioned however that an additional lesson learned in Chapter 4 with the need for full characterisation with rigorous controls for any design that proved promising in the lysate screens.

5.2.2. Design and assessment of FlAsH binding sites across Dps dimer-dimer interfaces

There are a number of considerations that must be made in designing binding sites across dimer-dimer interfaces in Dps. The Dps nanocage has tetrahedral symmetry, thus, unique interactions are repeated around the nanocage in a symmetry-related manner. The

dimer-dimer interfaces, while not located directly at symmetry axes are reliant on this symmetry nonetheless. Therefore there are a limited number of unique dimer-dimer interactions, but they are repeated multiple times in the assembled nanocage. In addition, the location of native cysteines must also be considered. Dps has a single cysteine close to the C-terminus of the monomer, well away from a dimer-dimer interface (see Appendix 5.2). While the single native cysteine will not impede this investigation, consideration must be taken when employing this approach with other proteins.

Dps was analysed to identify the major type of dimer-dimer interaction (Figure 5.2). This, the 'T-interaction', is formed between a dimer and a second dimer which resembles a "cap stone" on top of the first. Deeper analysis of this interface using the criteria defined above leads to only one area of potential interest (positions in red, Figure 5.2). This area is made up of the unstructured N-terminal tail from one monomer in one dimer and parts of both the C- and D-helix from a monomer in the other dimer. Thus, two complementary locations on opposing monomers were provided for possible mutations to cysteines to create a FlAsH binding site. These residues obey most of our design rules: each residue is located on a monomer that belongs to a different dimer and each unit is close in space. Through further examination of the crystal structure at this region, it is evident that some of the residues on the C-helix are buried and therefore inaccessible. The remaining residues are on the N-terminal tail (R18, N19, D20) and the D helix (D123, A126, N130). These six residues were the focus of our subsequent design strategy.

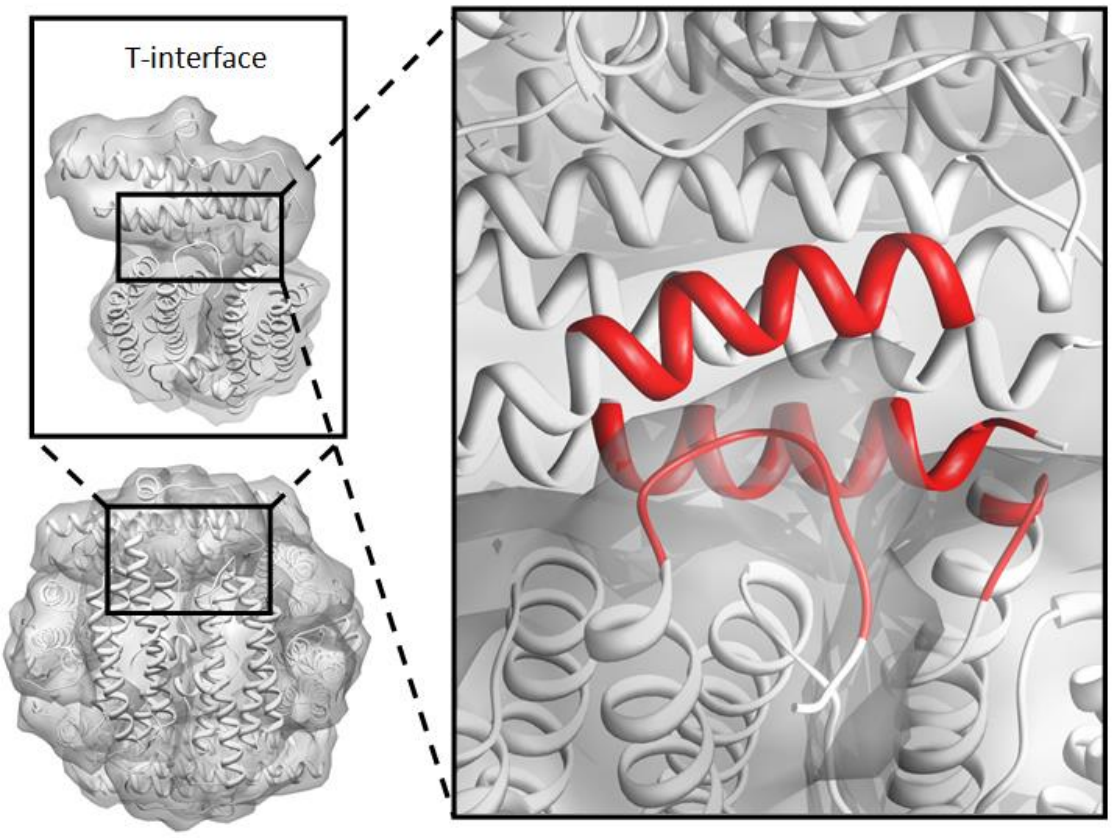


Figure 5.2 – Structural analysis of Dps dimer-dimer interfaces. (bottom left) Crystal structure (PDB:1DPS) showing a Dps ferritin nanocage viewed down a two-fold axes between monomers forming an anti-parallel dimer. (top) expansion of the T-interaction highlighting residues (red) that obey the design rules (see text). Protein figures made with Chimera[8].

It was unclear which four of these six residues were optimal; therefore, we decided to try all possible combinations. It should also be pointed out that because Dps is a highly symmetric homooligomer, each monomer is involved in two of these interactions at two different positions—once in the D-helix and the other near the N-terminal tail. Therefore, four cysteine mutations must be incorporated into each monomer to form the binding site in the assembled nanocage. Mutants were formed and examined (**DpsR18C-N19C-A126C-N130C, DpsN19C-D20C-A126C-N130C, DpsR18C-N19C-D123C-A126C, DpsR18C-D20C-D123C-A126C, DpsN19C-D20C-D123C-A126C, DpsR18C-D20C-A126C-N130C,** Figure 5.3) with some combinations omitted as they created binding sites too large for ideal FlAsH-EDT_2 geometry as defined by Schepartz and co-workers[9].

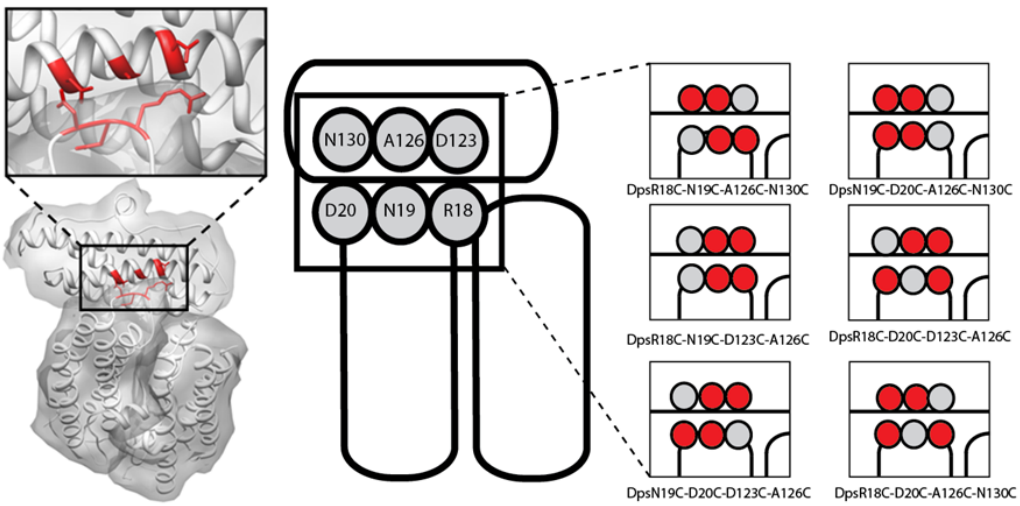


Figure 5.3 – Location and description of designed binding sites across Dps T-interfaces. (left) Dps T-interface with N-terminal tail R18, N19 and D20 and D-helix D123, A126, N130 amino acids highlighted (red). (Right) schematic of Dps T-interface expanded to show which amino acids (in red) were mutated into cysteines (red balls) from their native amino acid (grey) in each design (top left: **DpsR18C-N19C-A126C-N130C**. Top right: **DpsN19C-D20C-A126C-N130C**. Middle left: **DpsR18C-N19C-D123C-A126C**. Middle right: **DpsR18C-D20C-D123C-A126C**. Bottom left: **DpsN19C-D20C-D123C-A126C**. Bottom right: **DpsR18C-D20C-A126C-N130C**.)

Using wild type Dps with no additional cysteines as a negative control (**Dps**), and a well-characterised Bfr positive control from Chapter 4 (**BfrCCPGCC-N-term**), the designs were analysed in lysate conditions for their ability to bind to FLAsH-EDT_2. As in Chapter 4, lysate conditions were deemed an advantageous starting point as the proteins require no purification leading to their faster development, the speed provides the ability to assess more designs, and the complex conditions provide insight into how such designs might behave *in vivo*. While most of the designs showed low signal in these conditions, **DpsR18C-N19C-D123C-A126C** exhibited a fluorescence intensity of around 20% of control (Figure 5.4).

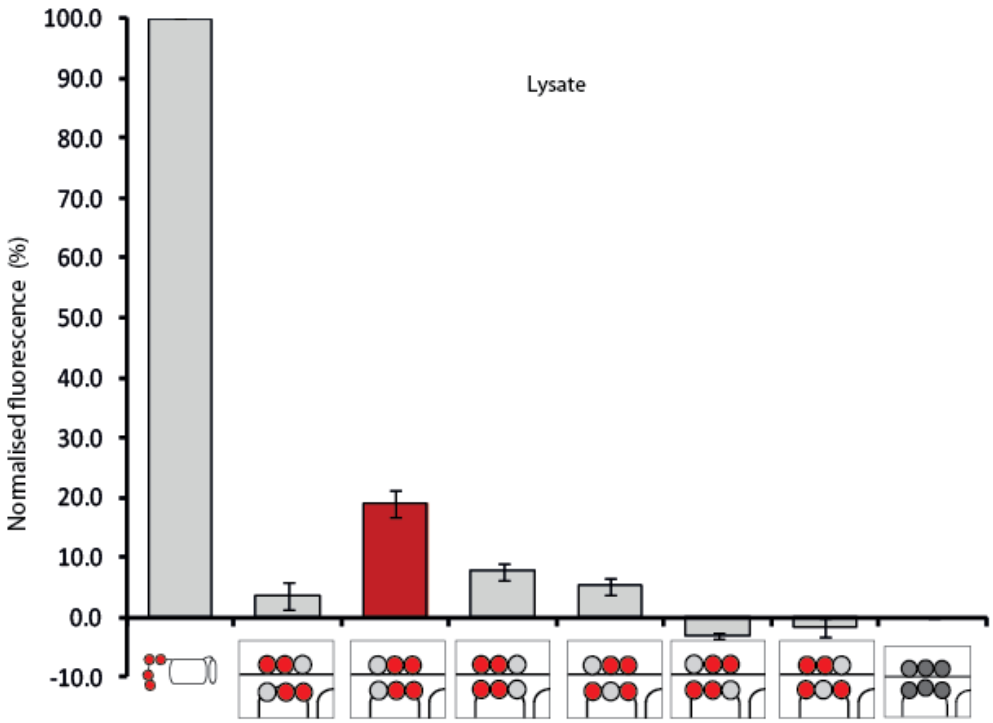


Figure 5.4 - FlAsH fluorescence of mutants with designed binding sites across Dps T-interfaces (see Figure 5.3 for schematic explanation) in lysates. The data is normalised to the positive and negative controls. All designs are 1 mg/ml in total protein concentration, in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) with 1 µM FlAsH, 3.5 mM TCEP, 1 mM EDT, 1 mM 2-ME. Signals are average of three separate protein expressions each repeated twice. Error bars are S.D.

Only one out of six designs showed significant fluorescence intensity above negative control, perhaps pointing to poor overall site geometry for FlAsH binding or to disruption of protein-protein interactions. This most promising design was denatured with the addition of 6 M guanidine·HCl to assess the effect of quaternary structure on the FlAsH signal following our procedure established in Chapter 4 (see Figure 5.5 A). The fluorescent signal decreased by nearly 100% suggesting that the fluorescent signal is dependent upon structured protein.

To confirm that the observed fluorescent signal observed in lysate experiments was arising from Dps nanocage assembly and not another species within the lysate sample, **DpsR18C-N19C-D123C-A126C** was purified and further characterised (see Appendix 5.4 for SDS-PAGE and Figure 5.5 for SEC). Unfortunately, during the purification process it became apparent that this protein was not in either the nanocage or dimer form (Figure 5.5 B). The SEC

analysis suggested **DpsR18C-N19C-D123C-A126C** had formed a much larger, possibly aggregated state. The aggregation of the protein, taken together with data from Chapter 4, begins to describe a narrative where possibly FlAsH-EDT_2 itself, the reducing conditions used, or the engineering of protein-protein interfaces leads to protein instabilities.

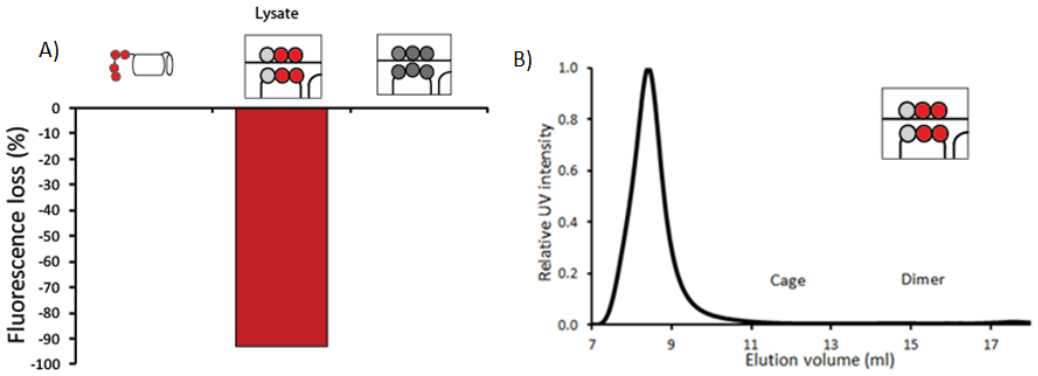


Figure 5.5 – (A) Assembly dependence of FlAsH fluorescence for the most successful mutant with a designed binding site across the Dps T-interface, **DpsR18C-N19C-D123C-A126C** (see Figure 5.3 for schematic explanation) in lysates. The data is presented as the percent change in fluorescence of the signal normalized to the positive and negative controls. All designs are 1 mg/ml in total protein concentration in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) with 1 µM FlAsH, 3.5 mM TCEP, 1 mM EDT, 1 mM 2-ME and 6 M guanidine·HCl (GuHCl). Signals are average of three separate protein expressions each repeated twice. (B) SEC of purified protein (0.5 mg/ml), **DpsR18C-N19C-D123C-A126C**. This data is the average of three repeats.

The philosophy of creating more precise FlAsH binding sites across a Dps protein-protein interface seems to result in similar problems to those observed with the terminus-focused Bfr designs described in Chapter 4. These problems could have been associated with redox conditions in the buffer as was suspected in Chapter 4 or with a non-ideal FlAsH binding site. Schepartz and co-workers determined that an ideal FlAsH binding site is a parallelogram with sides of 4-5 A and 6.8 A[9]. Analysis of the position of the four cysteines in **DpsR18C-N19C-D123C-A126C** suggested that they are close to the ideal geometry, although far from symmetric. (Figure 5.6) (It should be noted that this analysis did not perform any energy minimizations to determine the lowest energy conformers.) However, this does not account for the disruption of nanocage assembly. A third explanation may lie within the underlying philosophy of this design itself. As part of the justification for the terminus-based designs in

Chapter 4, we suggested that designing a binding site across a protein-protein interface may cause a disruption of a key interaction. It is possible that the mutation across the T-interface destroyed key protein-protein interfaces which resulted in a disruption of the nanocage. Therefore, we decided to pursue a terminus-focused approach.

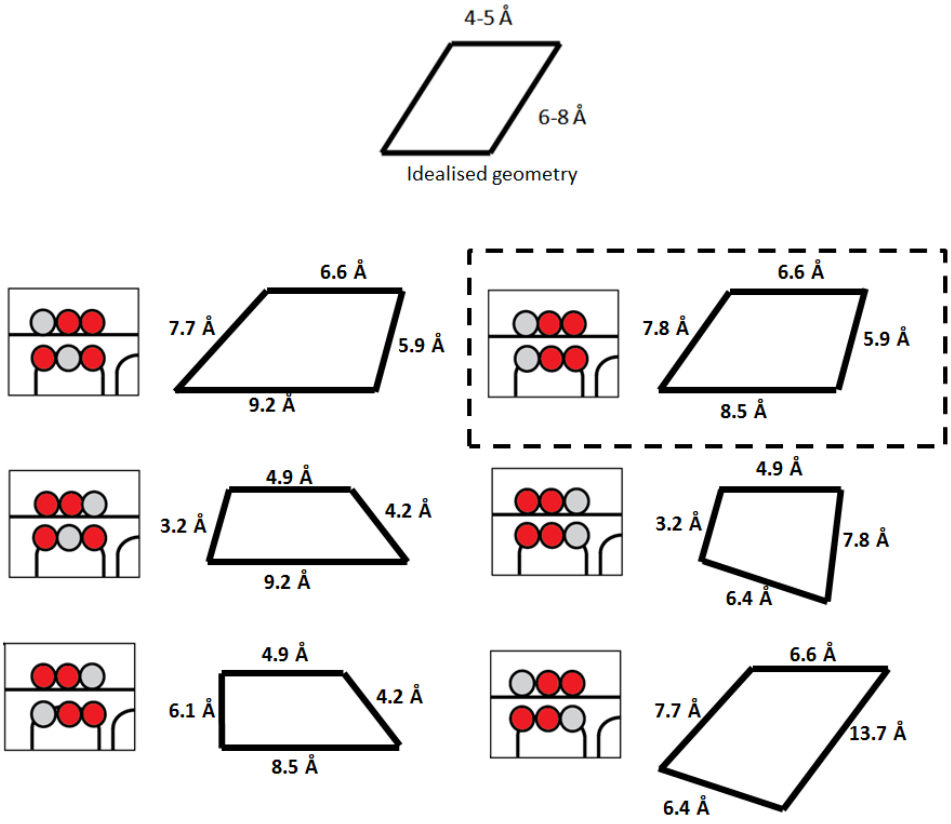


Figure 5.6 – Analysis of Dps T-interfaces (see Figure 5.3 for schematic explanation) FlAsH binding site geometry. (top) Optimal FlAsH binding site determined by Schepartz and coworkers, (bottom) distances between the cysteine sulfur atoms from each design determined from a crystal structure. The design (**DpsR18C-N19C-D123C-A126C**) that was the study of this investigation is highlighted (black dashed box). Crystal structure modifications made using Acceryl software[10] to place cysteines into the correct position for geometry analysis.

5.2.3. Design and assessment of FlAsH binding sites at Dps termini

While extending flexible termini to contain bipartite cysteine pairs (as seen in Chapter 4) provides a less precise design than engineering binding sites across a protein-protein interface, it could bypass any problems associated with protein-protein interaction disruption. The N-terminus of Dps is long (21 amino acids) and unstructured to the extent that the crystal structure has little electron density at the first residue making its exact location unknown. The

C-terminus however is more ordered. Therefore one possible strategy would be to design binding sites at the C-terminus of Dps that exploit the divergence of monomer termini in the antiparallel dimer, but which converge during nanocage formation. Thus, terminally appended cysteines could provide a sulfur rich area for a FlAsH binding site that only forms upon nanocage assembly and not in a dimer intermediate.

Designs were created by placing cysteine pairs at the Dps C-terminus (Figure 5.7). A series of mutants were generated to optimise the binding site. These designs differ in how the two cysteines are displayed by the monomer either based on linker length (**DpsCC**, **DpsGCC**, **DpsGGCC**, **DpsGGGCC**) or conformation (**DpsPAGCC**). It was thought that different conformational presentations might be necessary because of the lack of flexibility at the Dps C-terminus. Extending the C-terminus to display the peptide sequence CCPGCC which includes all four cysteines of an ideal binding site on a single monomer, resulted in a Dps positive control (**DpsCCPGCC**) which would require neither folding nor assembly to generate a FlAsH signal. Wild type Dps, which has no additional cysteines, was used as a negative control (**Dps**)**.**

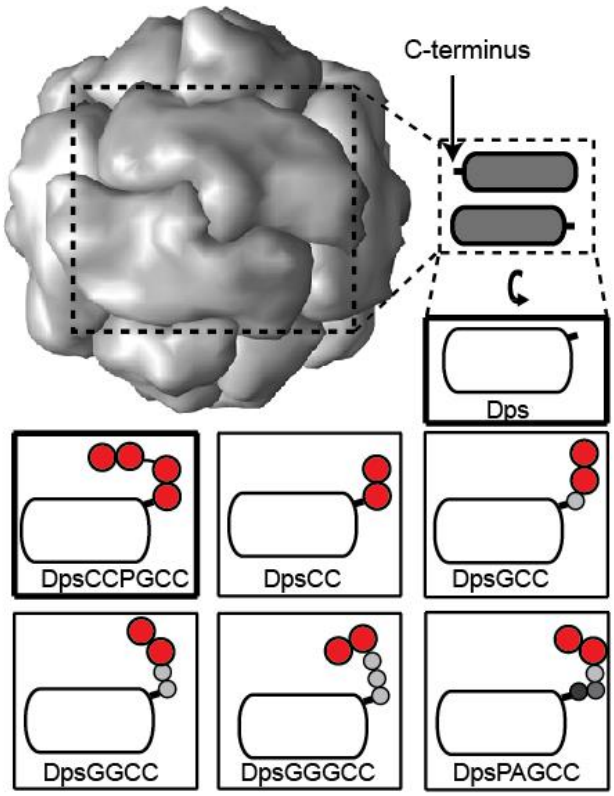


Figure 5.7 – Dps crystal structure (PDB: 1DPS) emphasizing a dimer subunit and divergent presentation of the C-termini along with schematics representing proteins used in this study. The negative control (**Dps**) has no appended C-terminal cysteines whereas the monomer of the positive control (**DpsCCPGCC**) displays a full binding site. The other proteins present a pair of cysteines with a variety of flexible and constrained linkers. Cysteine (red), glycine (light grey), alanine (dark grey), proline (black).

An oligomerization state assay with the most utility would be able to evaluate multiple samples rapidly. Therefore, instead of initially working with purified proteins, cellular lysates were used (as in Chapter 4 and above). In addition, the use of complex solutions allowed for the optimisation of this assay for selectivity. However, these samples were initially screened with a plate reader instead of a steady state fluorimeter as was used in Chapter 4. This plate reader exhibited far higher sensitivity than the steady state fluorimeter which permitted a decrease in both the fluorophore and protein concentrations (FlAsH-EDT_2 - 0.1 µM, protein - 0.1 mg/ml). Under these conditions, (Figure 5.8) the positive control, **DpsCCPGCC**, generated robust fluorescence with added FlAsH, and the negative control (**Dps**) gave nearly undetectable signal. Of the bipartite designs, **DpsCC** and **DpsGCC** had a very weak response whereas **DpsGGCC**, **DpsGGGCC,** and **DpsPAGCC** gave fluorescence that was between 15 and

35% of the positive control suggesting that a longer linker is ideal and one that is more rigid may be optimal.

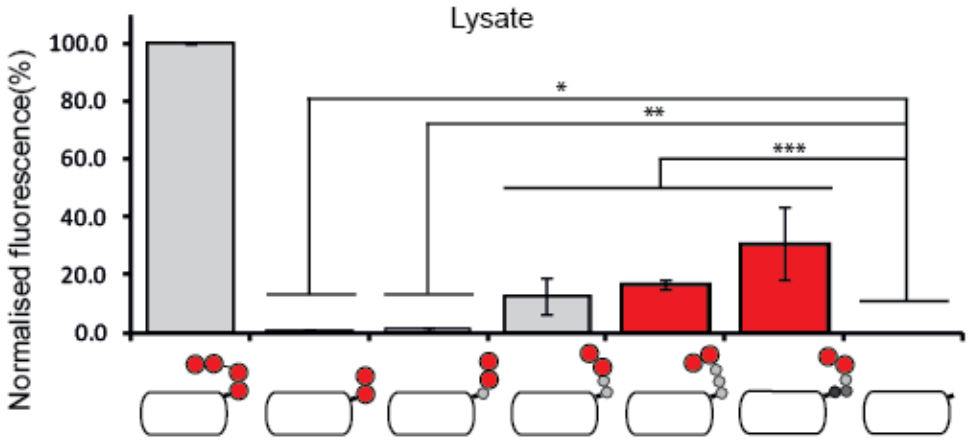


Figure 5.8 - FlAsH fluorescence of designs with bipartite FlAsH binding sites at the Dps C-terminus (see Figure 5.7 for schematic explanation) in lysates. The data is normalised to the positive and negative controls. All designs are 0.1 mg/ml total protein concentration, in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) with 0.1 µM FlAsH, 3.5 mM TCEP, 1 mM EDT, 1 mM 2-ME. The fluorescence is the average of three protein expressions each repeated twice. Error bars are S.D. * Two-tailed P-values = 0.4804. ** Two-tailed P-values = 0.020050. *** Two-tailed P-values = 0.0001.

To confirm that FlAsH-EDT_2 binding was indeed dependent on nanocage formation, the samples were reanalysed in denaturation conditions (Figure 5.9). This experiment was normalised differently to the other denaturing experiments in Chapter 4 and above. In this case we used an un-denatured positive control normalised to 0% change, to show the effect of denaturant on the controls. As expected the negative control, **Dps**, exhibited no change in fluorescence upon denaturation. The positive control, **DpsCCPGCC,** was also expected to have no change; however, it did generate a small, but significant loss which may be attributed to restricted access to the tetracysteine tag similar to that seen in Chapter 4. This alteration in experimental methodology was due to a better understanding of working with FlAsH because of the extended work performed in Chapter 4. As expected, the bipartite designs displayed a large loss in fluorescence upon denaturation with **DpsGGGCC** and **DpsPAGCC** exhibiting nearly a 100% loss in signal with added denaturant.

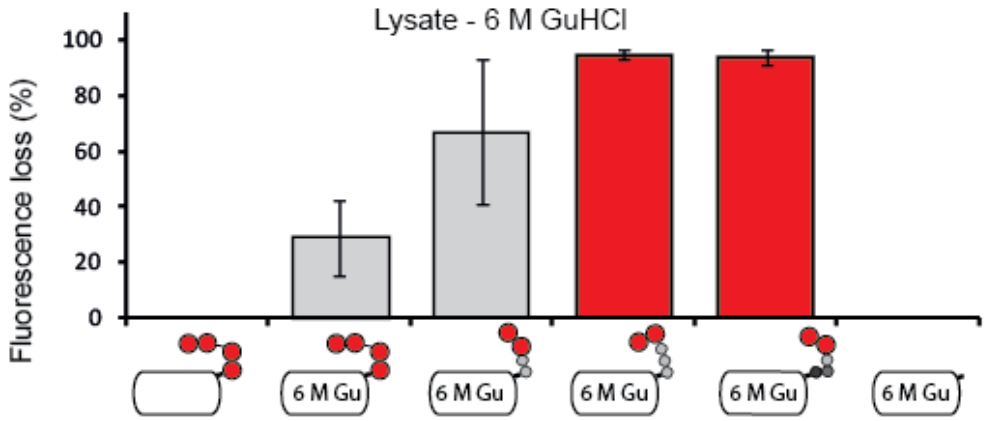


Figure 5.9 - The assembly dependence of FlAsH fluorescence for designs with bipartite FlAsH binding sites at the Dps C-terminus (see Figure 5.7 for schematic explanation) in lysates under denaturing conditions. The data is the change in the signal normalised to the positive and negative controls. All designs are 0.1 mg/ml in total protein concentration in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) with 0.1 µM FlAsH, 3.5 mM TCEP, 1 mM EDT, 1 mM 2-ME and 6 M guanidine·HCl (GuHCl). Data is the average of three separate protein expressions each repeated twice.

To verify these results, the leading proteins were expressed, purified and fully characterised (see Appendices 5.4, 5.5, 5.6, 5.7, 5.8, 5.10 and 5.5 for SDS-PAGE, mass spectrometry, TEM and CD and Figure 5.10 for SEC) where all proteins exhibited nanocage character in both TEM and SEC with the largest population being the nanocage oligomerization state.

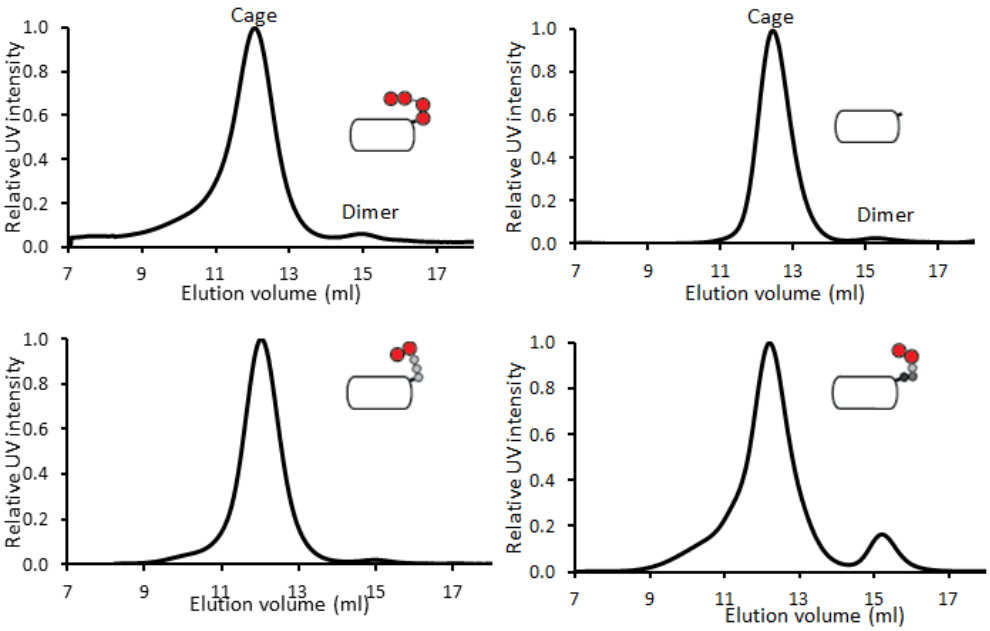


Figure 5.10 - Size exclusion chromatograms of purified designs with bipartite FlAsH binding sites at the Dps C-terminus monitored at 280 nm absorbance (protein) normalised to their highest intensity. Results for the UV are averages of three runs (0.5 ml injection of 0.5 mg/ml protein).

These proteins were subjected to the binding and subsequent denaturation experiments (Figures 5.11 and 5.12). The controls behaved as expected with the exception of the positive control that, consistent with the data in lysates (Figure 5.9) lost a small amount of signal upon denaturation. The bipartite designs, **DpsGGGCC** and **DpsPAGCC,** displayed strong, oligomerization dependent, binding to FlAsH with the latter protein generating ~60% of control.

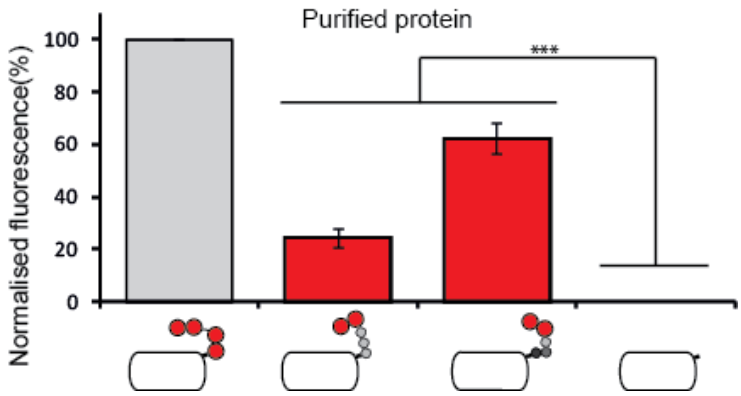


Figure 5.11 - FlAsH fluorescence of purified designs with bipartite FlAsH binding sites at the Dps C-terminus (see Figure 5.7 for schematic explanation). The data is normalised to the positive and negative controls. All designs are 0.1 mg/ml protein, in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) with 0.1 µM FlAsH, 3.5 mM TCEP, 1 mM EDT, 1 mM 2-ME, averaged from 8 replicates. Error bars are S.D. *** Two-tailed P-values = 0.0001.

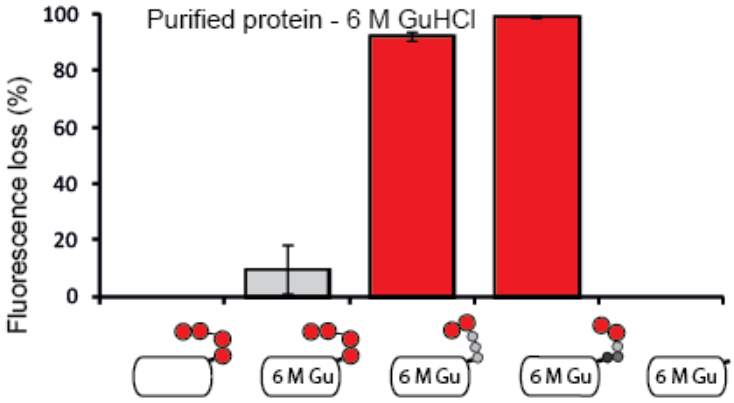


Figure 5.12 – The assembly dependence of FlAsH fluorescence for purified designs with bipartite FlAsH binding sites at the Dps C-terminus (see Figure 5.7 for schematic explanation) in denaturing conditions. The data is the change in the signal normalised to positive and negative controls. All proteins are 0.1 mg/ml protein concentration, in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) with 0.1 µM FlAsH, 3.5 mM TCEP, 1 mM EDT, 1 mM 2-ME and 6 M guanidine·HCl (GuHCl). Data is the average from 8 replicates. Error bars are S.D.

The steady state fluorescence experiments, coupled with denaturation, suggest that split cysteine bipartite technology with the Dps C-terminus-focused designs was successful. The strong signals seen in lysate and purified samples that greatly decrease when denatured, as well as the presence of nanocage structures in both SEC and TEM, provide initial indication that these designs could be good candidates for a broader system to follow protein nanocage assembly. However, our experience has shown us that they need to be more rigorously characterised.

5.2.4. Identification of fluorescent oligomerization state for designs with bipartite FlAsH binding sites at the Dps C-terminus via SEC-FPR

The oligomerization of the designs with bipartite FlAsH binding sites at the Dps C-terminus is much more defined in favour of the nanocage state than our other design strategies (above and Chapter 4). Very little dimer and undefined aggregate was evident. However, it is essential to determine which species is responsible for fluorescence. Therefore SEC-FPR which combines the SEC absorbance detector with a fluorescence plate reader, was employed (Figure 5.13). If the binding sites were successfully designed to appear only during the formation of the nanocage, then the fluorescent peaks should correspond to only that state. The negative control, wild type **Dps**, generated a single peak from the nanocage and nearly undetectable dimer when monitored at 280 nm (Figure 5.13b) as expected. Also as expected, no nonspecific interactions were observed in the FlAsH channel. The positive control, **DpsCCPGCC**, generated only a nanocage and a small dimer peak, both of which were positive in the fluorescence channel, again as expected (Figure 5.13a). The proteins with the bipartite, C-terminal fusions, **DpsPAGCC** and **DpsGGGCC**, both showed a nanocage and a small dimer peak in the 280 nm channel, however, fluorescence only elutes with the nanocage, suggesting that FlAsH is specific for the nanocage (Figures 5.13c and d). It should be noted that **DpsPAGCC** did aggregate slightly in these conditions, but the fact that the aggregate peak also had no fluorescence further demonstrates the robustness of this approach compared to that

of Bfr. As a further analysis of specificity, the same technique was used to assess clarified lysates for **DpsPAGCC** and **DpsGGGCC** (Figures 5.13e and f). Again, fluorescence elutes at a volume consistent with the nanocage state. Taken together these data strongly suggest that not only are the designed binding sites forming upon oligomerization, they form specifically only upon nanocage formation.

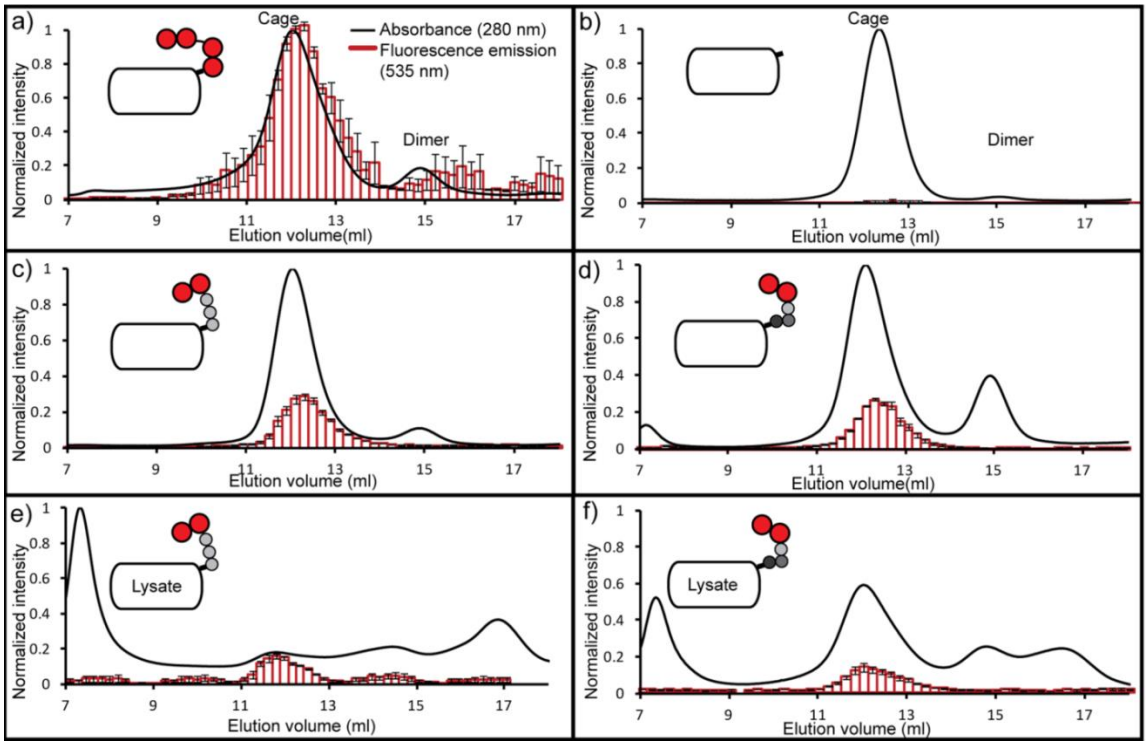


Figure 5.13 - Identification of fluorescent oligomerization state for designs with bipartite FlAsH binding sites at the Dps C-terminus via SEC-FPR (see Figure 5.7 for schematic explanation). Size exclusion chromatograms were monitored at absorbance 280 nm (protein). Fluorescence samples were taken every 200 µl and analysed using a 510 ± 5 nm excitation and a 530 ± 5 nm emission filter set. Each run injected 0.5 ml of 0.5 mg/ml purified protein or total protein (lysate samples) in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) with 0.4 µM FlAsH, 3.5 mM TCEP, 1 mM EDT, 1 mM 2-ME. The data is the average of three runs per protein. Data is normalised using the highest fluorescence signal from the positive control as 100% for all samples and the lowest of each as 0%. Error bars are S.D.

5.2.5. Medium throughput screen of nanocage stability conditions using DpsPAGCC

With the design of a bipartite FlAsH binding site at the Dps C-terminus proven to be robust, we set out to fulfill one of our more advanced goals to utilise **DpsPAGCC** in an assay with higher throughput. As a proof of principle, we expanded the assay to a 96-well plate format and screened multiple buffer conditions for their ability to stabilise the nanocage state.

Both pH and denaturant were screened simultaneously (Figures 5.14 and 5.15). As expected, there was little significant change in fluorescent signals from the controls **DpsCCPGCC** for any condition although it should be noted that these provide an indication of the pH dependence of FlAsH binding in these buffer conditions. For **DpsPAGCC,** a wide range of signals was observed across the window of conditions. Most interestingly, the protein seemed most stable below pH 7.0. This is somewhat surprising considering that most published accounts of research on this protein describe experiments performed at pH 7.8[5, 11]. Thus it was affirming that even this simple proof of principle experiment could make a contribution to the field. More broadly, it emphasizes the utility of a rapid, direct, and less reagent intensive assay for protein nanocage formation. It not only underlines the strength of our approach, but this increased throughput can lead to useful experiments (with potentially unexpected results) that may not have been run due to the exigent nature of traditional techniques.

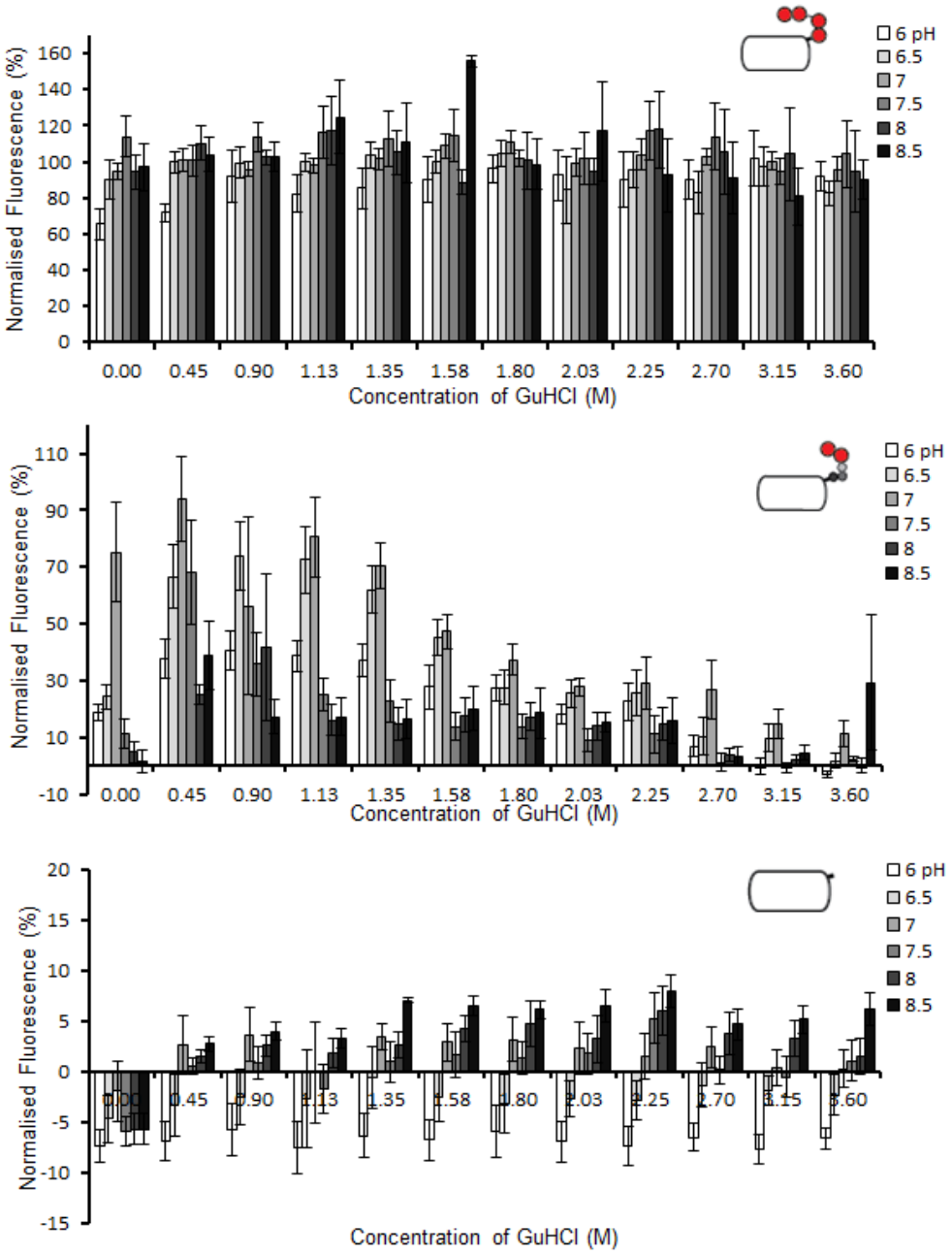


Figure 5.14 – Medium throughput screening of buffer conditions for protein nanocage stability in a 96-well format. Normalised fluorescence data for pH vs GuHCl for designs with bipartite FlAsH binding sites at the Dps C-terminus (see Figure 5.7 for schematic explanation). This data is presented more clearly in Figure 5.15. Data normalised to the controls with the average response from **DpsCCPGCC** being 100% and the average response of **Dps** being 0%. All designs were 0.1 mg/ml total protein concentration in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) with 0.1 µM FlAsH, 3.5 mM TCEP, 1 mM EDT, 1 mM 2-ME. Data is averaged from 3 replicates. Error bars are S.D.

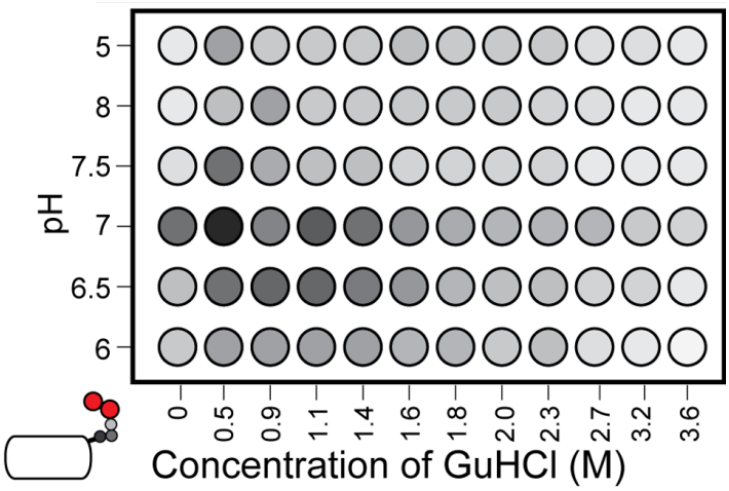


Figure 5.15 - Medium throughput screening of buffer conditions for protein nanocage stability in 96-well format. Normalised fluorescence intensity for **DpsPAGCC** as a function of both pH and Gu·HCl. This data is a visualization of the data presented in Figure 5.14. Fluorescence intensity is grey-scaled so that black is most intense and white is least. The protein is 0.1 mg/ml in 0.1 µM FlAsH, 3.5 mM TCEP, 1 mM EDT, 1 mM 2-ME. The data is the average of 3 replicates.

To explore the scope of this technology, we were curious to determine whether it is possible to disassemble the Dps nanocage without denaturing it. Along with useful knowledge that may aid future research with these proteins, it may also point towards the mechanism of assembly. Moreover, it could lend insight into how to disassemble and reassemble these proteins with the intention of loading them with "cargo" for eventual delivery applications as discussed in Chapter 2. Therefore, a similar 96-well plate denaturation experiment was performed with **DpsPAGCC** at pH 7.8 as described above (Figure 5.16) and this data was overlaid with a CD denaturation experiment (Figure 5.17). Circular dichroism monitors secondary structure whereas our technique is specific for the nanocage quaternary structure. The quaternary structure assay (Figure 5.17) performed as expected with the controls having very little change with **DpsPAGCC** disassembling at 1.4 M denaturant.

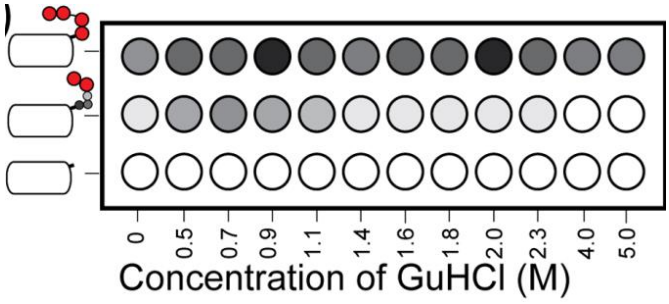


Figure 5.16 – Monitoring disassembly of **DpsPAGCC** in the 96-well format. Fluorescence intensity is grey-scaled so that black is most intense and white is least. Data normalised to controls. This data was used to create the plot in Figure 5.17. All proteins are 0.1 mg/ml in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) with 0.1 µM FlAsH, 3.5 mM TCEP, 1 mM EDT, 1 mM 2-ME. Data is averaged from 3 replicates. Error bars are S.D.

The secondary structure of **DpsPAGCC** was monitored with respect to guanidine by its ability to differentially absorb 222 nm (a major absorbance for the alpha helix secondary structure) plane polarized light (Figure 5.17 and Appendix 5.9 for CD scans). The protein loses the majority of its secondary structure after the addition of 2.2 M guanidine·HCl. The difference between the concentration required to disassemble the protein and to totally denature it could present a window where the disassembled-but-folded nanocage is stable. Our ability to rapidly perform this analysis further emphasizes the strength of our approach.

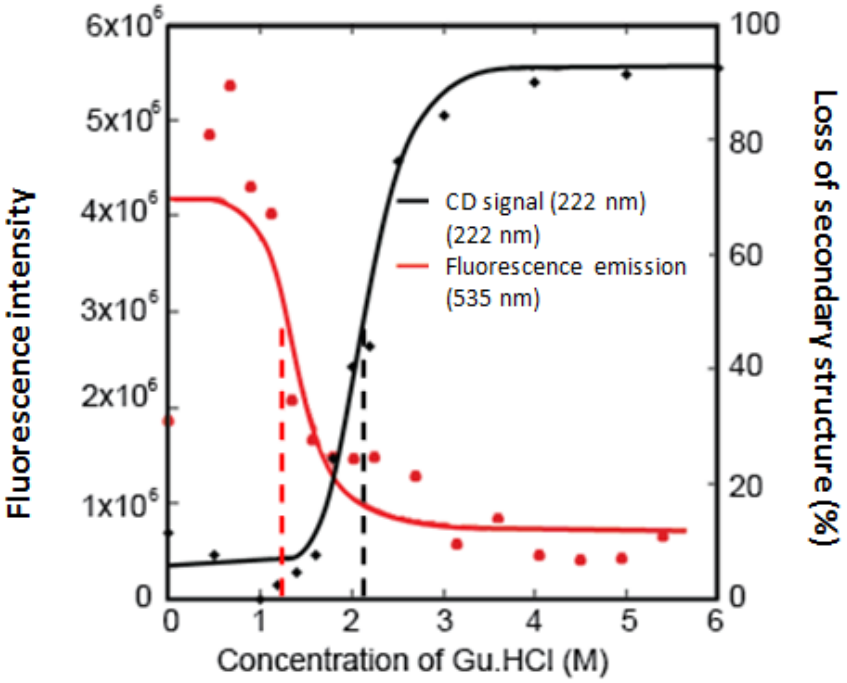


Figure 5.17 – Determination of stability window for disassembled-but-folded protein nanocage. Overlay of **DpsPAGCC** FlAsH fluorescence (510 ± 5 nm excitation and a 530 ± 5 nm emission filter set, red line) and loss of protein secondary structure, as monitored by CD (222 nm black line), with respect to guanidine·HCl concentration. Fluorescence data is plotted from data presented in Figure 5.16 from 3 repeats, error bars not shown. Fluorescence experiment is 0.1 mg/ml total protein concentration, in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) with 0.1 µM FlAsH, 3.5 mM TCEP, 1 mM EDT, 1 mM 2-ME, averaged from 3 replicates. Circular dichorism is 0.1 mg/ml protein concentration in FlAsH buffer with taken from a single scan of each condition. Fluorescence fitting made using Kaleidagraph (Synergy) calculated to 1.39 M Gu·HCl and CD fitting made also using Kaleidagraph (Synergy) calculated to 2.15 M Gu·HCl.

5.2.6. Medium throughput screen of nanocage stability in nanoparticle genesis conditions

Our technique, in addition to helping to explore protein biochemistry and biophysics, could aid in the manipulation and engineering of proteins for novel applications. An issue that arose when trying to expand hybrid nanoparticle synthesis technology to bacterial ferritins (see Chapter 2) was protein stability and the extended time it took to determine whether the proteins had survived the multi-step process. The slow characterisation time lead to a reduction in the number of experiments that could be attempted. Therefore we set out to determine whether it is possible to apply the medium throughput assay for nanocage stability to the screening of conditions that would be typical for the production of gold nanoparticles[12] and to isolate potential problems.

Different concentrations of reagents for the generation of gold nanoparticles inside ferritins ($HAuCl_4$, $NaBH_4$, $NaCNBH_3$, trisodium citrate, ascorbic acid) were screened individually in buffer or water for their ability to disrupt **DpsPAGCC** assembly (Figure 5.18). Sodium cyanoborohydride and trisodium citrate showed stable protein nanocages across the range of concentrations in buffer. Interestingly, in water, citrate while having very little concentration dependence showed increased stability over protein alone. Furthermore, cyanoborohydride caused a steady decline in assembly with respect to concentration. These results might be an extension of the pH trend seen in the experiment presented in Figure 5.15 where more acidic conditions were more stable than basic ones although it should be noted that ascorbic acid induced strong destabilisation in unbuffered conditions. The protein assembled state is quite sensitive to chloroauric acid solution, but buffered conditions can extend the stability into a rather flexible window, which confirms the trend seen in Chapter 2. Curiously, sodium borohydride, which appears to be more stabilizing at low concentrations in water over that of buffered conditions however, at higher concentrations, the stability falls off rapidly in both water and buffer. The comparison between sodium borohydride and sodium cyanoborohydride is striking and suggests that it was a good choice to switch to sodium cyanoborohydride in Chapter 2.

This example further demonstrates the practical utility of our method. It permits the rapid analysis of reagents and their usable concentration ranges. While only limited reagents were used in this study, it could easily be extended to examine other reagents or reagent combinations in a rapid and parallel manner. The straightforward “enabling” aspect of this technique is underlined when reflecting, perhaps painfully; at least to the author of this thesis, on the impact it, and these data, could have made to facilitate problem solving in Chapter 2.

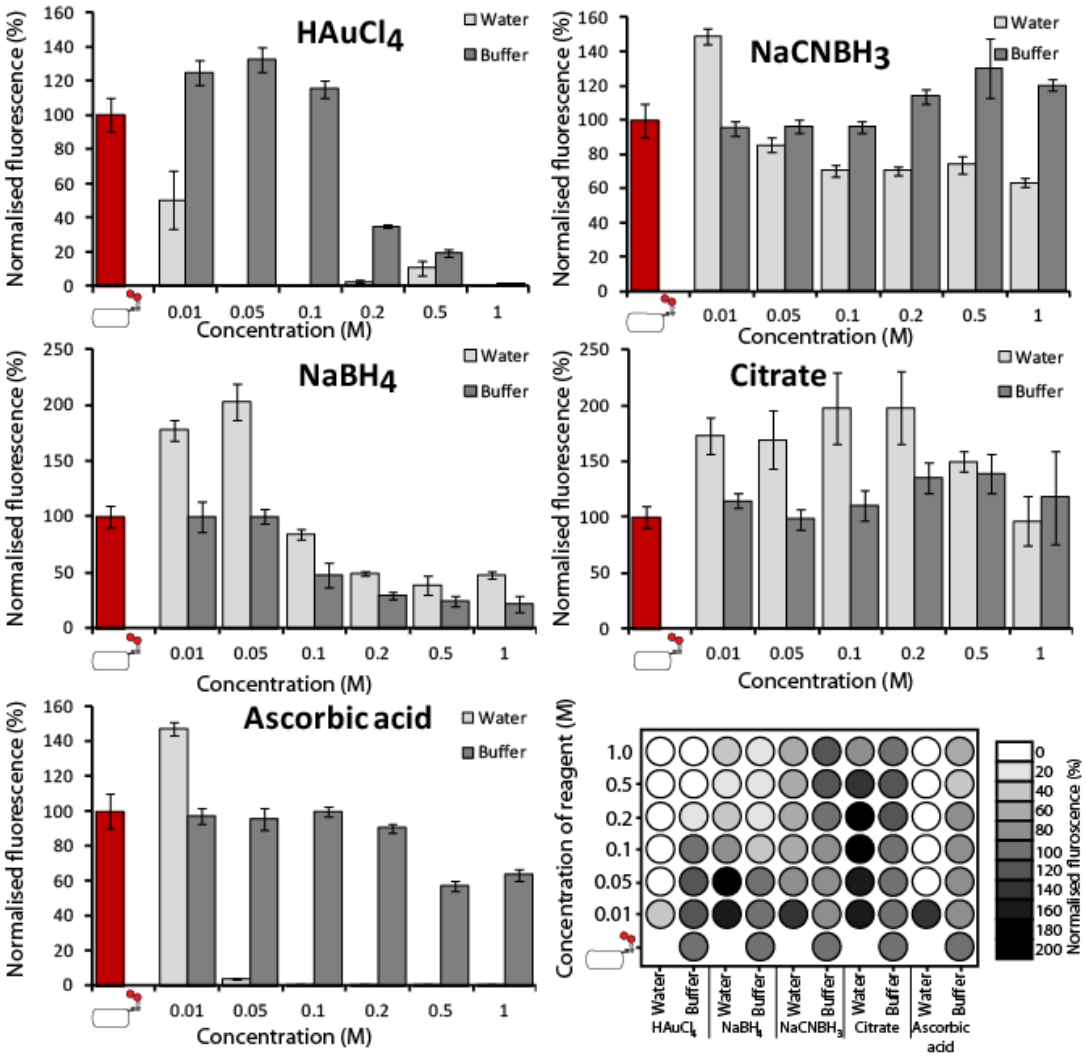


Figure 5.18 – Medium throughput screen of **DpsPAGCC** stability to nanoparticle genesis reagents (top left) chloroauric acid (top right) sodium cyanoborohydride (middle left) sodium borohydride (middle right) trisodium citrate (bottom left) ascorbic acid. (Bottom right) 96 well plate format with greyscale corresponding to fluorescence intensity. The data is normalised to **DpsPAGCC** in buffer with no added reagents. Each reagent was analysed both as a solution in water (light grey) and in Flash buffer (dark grey). Proteins were 0.1 mg/ml in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) with 0.1 µM FlAsH, 3.5 mM TCEP, 1 mM EDT, 1 mM 2-ME. All data is averaged from 3 replicates. Error bars are S.D.

5.3. Conclusion

In summary, the goal of developing a method for the direct detection of protein nanocage assembly was refocused onto the mini-ferritin Dps, and in this case, the goal was achieved. With lessons learned from Chapter 4, design rules were applied including a requirement for any binding site to be on the exterior of the nanocage, and two specific philosophies were pursued. One philosophy involved designs spanning key protein-protein interfaces. Mutations straddling the T-interface were positioned on the D-helix and the N-terminal loop of Dps. Although these designs fulfilled the design rules and geometric requirements for FlAsH binding, and some exhibited denaturation-dependent fluorescence signals in lysates, no folded protein was observed.

A second philosophy was pursued similar to that used in Chapter 4 where cysteine pairs were fused to the C-terminus of Dps monomers. Multiple designs were explored that differed in the linker length and conformational presentation of the cysteine pair. The longest and most rigid of these designs showed initial fluorescent signals in lysates that diminished upon denaturation. Moreover, this trend was repeated in purified proteins, and nanocage architectures were observed in both SEC and TEM. Furthermore, SEC-FPR demonstrated that the fluorescence signal co-eluted only with the nanocage confirming the ability for this assay to selectivity detect nanocage formation over other oligomerization states. With this design, the assay was expanded to a medium throughput screen, to discover the ideal conditions for Dps nanocage stability as well as to investigate destabilising conditions when attempting to form encapsulated nanoparticles.

While this chapter describes a large step forward the goal of developing a rapid and direct strategy to assess protein nanocage formation in complex solutions, a question is raised as to why this strategy was successful for the Dps model system but not for Bfr, especially considering the similarities between their secondary and tertiary folds (see Chapter 3). The

origins of this should be thoroughly scrutinized if this technology is to be expanded to other nanocages. While the thermal and chemical denaturation stabilities for these proteins do not differ greatly, behavioural differences could arise from differences in structural energetics[3, 13]. Bfr was initially chosen as a model system due to the fact that it forms mixtures of nanocage and dimer in solution; the thought being that having two states, and mutants that provide different ratios of these states, would be an advantage towards assessing oligomerization state selectivity and thus higher stringency in assay development. Because this was later seen as an unnecessary complication, Dps was chosen for the second attempt due to the fact that the dimer is much less prevalent. Although the FlAsH methodology has been successful for Dps, and could also be used for other nanocages, the issues experienced with Bfr could limit its applications to other proteins. However, the characterisation methods in Chapters 4 and 5 provide rigorously controlled approaches for the detection of successful designs over non-successful ones. Thus, our success, while perhaps limited in scope is still a step forward for both split tetracysteine technology, showing that is can work in large complex multimeric proteins, as well as protein-protein interaction research, presenting a new method that could be used to aid future investigations in this area.

While the success of this method could help in speeding and providing direct information into assembly *in vitro*, this method is still limited to a medium throughput. To use it to greatly aid nanocage and quaternary structure research, further expansion of the throughput of the method is essential.

5.4. Methods

5.4.1. Cloning of Dps designs

5.4.1.1. Cloning of all Dps designs for screening in protein lysates (Dps, DpsCC, DpsGCC, DpsGGCC, DpsGGGCC, DpsPAGCC, DpsCCPGCC, DpsR18C-N19C-A126C-N130C, DpsN19C-D20C-A126C-N130C, DpsR18C-N19C-D123C-A126C, DpsR18C-D20C-D123C-A126C, DpsN19C-D20C-D123C-A126C, DpsR18C-D20C-A126C-N130C)

The WT Dps gene[4] was amplified with primers providing restriction sites. The PCR reaction (50 µl total volume) was a solution of reaction buffer (Fermentas, 5 µl of 10xPfu buffer), dNTP mix (Fermentas, 2 µl of a solution containing dATP, dTTP, dGTP, and dCTP 2 mM each), forward and reverse primers (400 ng each, Suprenom, Singapore, see Appendix 5.1 for primers), template (100 ng) and Pfu polymerase (Fermentas, 2 µl of 2.5 U/ µl) which was subjected to an initial melting step at 95 °C for 30 s, followed by 30 cycles of amplification (95 °C for 30 s, 55 °C for 45 s and 72 °C for 1 min) followed by 72 °C for 7 min. The resulting PCR product was isolated by gel purification (Promega). The PCR product (300 ng) was digested (Tango buffer, Fermentas, 10 µl of 2x buffer), Nde1 (Fermentas, 1.5 µl of 20 U/ µl), Xho1 (Fermentas, 1.5 µl of 20 U/ µl), 50 µl total volume, 37 °C, 3 hours). The plasmid pET-22b (Novagen) was treated in the same way, followed by ligation of the insert with T4 ligase buffer (Fermentas, 2 µl, 10x buffer), digested insert (30 ng), digested plasmid (80 ng), T4 ligase (Fermentas, 0.2 µL of 20 U/µl), 20 µl total volume (37 °C, 2 hours). The constructs (2 µl) were transformed (Novablue, Novagen) and the resulting colonies were assessed by colony PCR, a solution of Gotaq buffer (Promega, 5 µl of 5x), dNTP mix (Promega, 2 µl of a solution containing dATP, dTTP, dGTP, and dCTP at 2 mM each), T7 promoter and terminator primers (150 ng each, Eurofins MWG), Gotaq DNA polymerase (Fermentas, 0.25 µl of 5 U/ µl), colony suspension (10 µl of a 50 µl total colony suspension in deionised water) and deionised water (5 µl), was subjected to an initial melting step (95 °C for 10 min) followed by 30 cycles of amplification, ((95 °C for 30 s, 59 °C for 30 s and 72 °C for 30 s) followed by 72 °C for 7 min).

Clones demonstrating a PCR product of the predicted size were subjected to miniprep (Invitrogen) and the resulting purified plasmids were sequenced (see Sequence Appendix 5 for sequences and Appendix 5.1 for primers used).

Using the newly constructed Dps in pET-22b as a template, site directed mutagenesis was performed to create all other designs (Quikchange, Stratagene). The reaction mixture containing Pfu buffer (Promega, 2 µl of 10x), dNTP mix (Promega, 2 µl of a solution containing dATP, dTTP, dGTP, and dCTP at 2 mM each), forward and reverse primers (125 ng each, Suprenom, Singapore, see Appendix 5.1 for primers used), the dsDNA template (Dps pET-22b, 100 ng) and Pfu polymerase (Promega, 1 µl of 3 U/ µl) in a total volume of 25 µl was subjected to melting (95 °C for 5 min), followed by 20 cycles of amplification (95 °C for 1 min, 55 °C for 1 min and 68 °C for 6 min) followed by 70 °C for 10 min. The reaction was then digested with Dpn1 (NEB buffer 4 (1 µl of 10x), amplification product (5 µl), deionized water (3.5 µl), Dpn1 (NEB, 0.7 µl of 10 U/ µl), 37 °C, 2h). The resulting solution was transformed (XL-1 blue, Novagen) and the subsequent colonies subjected to miniprep (Promega, Wizard plus SV) and sequenced (see Sequence Appendix 5 for sequences and Appendix 5.1 for primers used).

5.4.1.2. Cloning of all designs for screening as purified proteins (Dps, DpsCCPGCC, DpsGGGCC and DpsPAGCC)

In order to express the proteins for isolation and purification, the genes were placed into a vector that provided affinity tags. The constructs were amplified from the mutated pET-22b plasmid with primers providing ligation independent cloning (LIC) sites. The PCR solution Pfu reaction buffer (Promega, 5 µl of 10x), dNTP mix (Fermentas, 2 µl of a solution containing dATP, dTTP, dGTP, and dCTP at 2 mM each), forward and reverse primers (400 ng each, Integrated DNA technologies, see Appendix 5.1 for primers used), the template (100 ng) and Pfu polymerase (Promega, 2 µl of 2.5 U/ µl) in 50 µl total volume was subjected to an initial melting step (95 °C for 30 s), followed by 30 cycles of amplification (95 °C for 30 s, 55 °C for 45

s and 72 °C for 1 min) followed by 72 °C for 7 min. The resulting PCR product was isolated by gel purification (Qiagen). The PCR product was then treated with T4 polymerase in order to create the required complementary overhangs (NEB buffer 2, 2 µl of 10x buffer), dATP (NEB, 2 µl of 25 mM), DTT (Sigma, 1 µl of 100 mM), BSA (NEB, 0.2 µl of 100x), PCR product (0.3 pmol) and T4 DNA polymerase (NEB, 0.6 µl of 10 U/µl) in a total volume of 20 µl, 30 min, 22 °C). The vector, pET-32b (Novagen) was treated in the same way. The insert was annealed to open plasmid (1 µl of the T4 Pol product, 0.5 µl of plasmid solution (50 ng/µl) by incubating (room temperature, 30 min) followed by a second incubation (room temperature, 30 min) with an addition of EDTA (1 µl of 100 mM). The constructs (2.5 µl) were transformed (XL-1 Blue, Novagen) and the resulting colonies were assessed by colony PCR. A solution of Gotaq reaction buffer (5 µl of 5x), dNTP mix (Promega, 2 µl of a solution containing dATP, dTTP, dGTP, and dCTP at 2 mM each), T7 promoter and terminator primers (150 ng each, Eurofins MWG), Gotaq DNA polymerase (Promega, 0.25 µl of 5 U/ µl), colony suspension (10 µl of a 50 µl total colony suspension in deionised water) and deionised water (5 µl), was subjected to an initial melting step (95 °C for 10 min) followed by amplification 30 cycles (95 °C for 30 s, 59 °C for 30 s and 72 °C for 30 s followed by 72 °C for 7 min). Clones demonstrating a PCR product of the predicted size were subjected to miniprep (Promega) and the resulting purified plasmid was sequenced (see Sequence Appendix 5 for sequences and Appendix 5.1 for primers used).

5.4.2. FlAsH Binding in Lysates.

T-interface Dps designs were analysed for their ability to bind to FlAsH in lysates as described in **4.4.2**.

The pET-22b expression vectors containing Dps designs with varying C-terminal linkers, were transformed into Rosetta *E.coli* cells (Novagen) and plated on LB agar plates (50 µl/ml of carbenicillin and 34 µl/ml chloramphenicol). Selected colonies then were grown in LB (1 ml with 50 µl/ml of carbenicillin, 37 °C, overnight) as a pre-culture which was then added to LB (100 ml) and grown (37 °C) until an $O.D_{600}$ of 0.6. Protein expression was then induced by the

addition of IPTG (50 µl of a 1 M stock) and the culture was further incubated (3 h, 30 °C). The cells were isolated by centrifugation (4,000 rpm, 15 min at 4 °C). The cell pellet was resuspended in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) and sonicated (Misonix, Ultrasonic cell disruptor, pulsed 5 s on 5 s off for 5 min). The protein solution was clarified by centrifugation (15,000 rpm, 45 min, 4 °C) and then filtered (Sartorius, 0.2 µm).

The protein concentration was determined (BCA, Novagen) and the cell lysate was diluted to 1 mg/ml with FlAsH buffer. To insure that all the 1 mg/ml samples had similar amounts of the desired protein, each was analysed by SDS PAGE (see appendix 5.3). Each fluorescence experiment contained protein lysate (200 µl, 0.1 mg/ml) in FlAsH buffer, TCEP (Sigma, final concentration of 3.5 mM), EDT (Sigma, final concentration of 1 mM) and 2-ME (Sigma, final concentration of 1 mM) were added and the solution was incubated (room temperature, 2 h) followed by the addition of FlAsH-EDT_2 (Invitrogen, final concentration of 0.1 µM) followed by a further incubation (room temperature, 2 h) in the dark. Each lysate sample was analysed in a black Corning 96 well plate in a PerkinElmer Envision 2101 multilabel plate reader, with each design being expressed three times and each expression being examined in 6 different wells and re-read twice (Ex filter 485 nm bandwidth 14 nm, Em filter 535 nm bandwidth 25 nm). For the denatured experiments, the above was repeated but with the addition of 6 M guanidine·HCl and incubated for 2 hours prior to the addition of TCEP, EDT and 2-ME, with protein concentration remaining the same as in previous un-denatured experiments.

5.4.3. FlAsH Binding with Purified Proteins.

The pET-32b vectors containing the Dps variants were transformed into Rosetta *E.coli* cells (Novagen) and plated on LB agar plates (50 µl/ml of carbenicillin and 34 µl/ml of chloramphenicol). Selected colonies were then grown in LB (5 ml, 37 °C, overnight) as a pre-culture which was added to LB (500 ml) and grown (37 °C) until an $O.D_{600}$ of 0.6. Protein

expression was then induced by the addition of IPTG (250 µl of a 1 M stock) and the cultures were further incubated (3 h, 30 °C). The cells were isolated by centrifugation (4,000 rpm, 20 min, 4 °C). The cell pellet was resuspended in lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 40 mM Imidazole, 1 mM EDTA, pH 8). Cellytic (10x, Sigma) was added and the solution was incubated (20 min, on ice) and then sonicated (Misonix, ultrasonic cell distruptor, pulsed 5 s on, 5 s off for 5 min). The protein solution was clarified by centrifugation (15,000 rpm, 45 min at 4 °C) and then filtered (Sartorius, 0.2 µm). The protein was purified via affinity purification (GE, Histrap FF, 5 ml, (wash buffer-40 mM Imidazole, 50 mM NaH_2PO_4, 300 mM NaCl, pH 7.4), (elution buffer-500 mM Imidazole, 50 mM NaH_2PO_4, 300 mM NaCl, pH 7.4)). Enteriokinase digestion (NEB 2 µg/ml) was performed to cleave off the peptide tag from the protein of interest followed by a second Histrap (GE, Histrap HP, 5 ml, (wash buffer-50 mM NaH_2PO_4, 300 mM NaCl, pH 7.4), (elution buffer-500 mM Imidazole, 50 mM NaH_2PO_4, 300 mM NaCl, pH 7.4)) to remove the tag from solution. The protein solution was further purified by size exclusion chromatography (GE Hiload 16/60 Superdex, running buffer- FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8).

Each purified protein was analysed by incubating the protein (200 µl of 0.1 mg/ml) in FlAsH buffer with TCEP (Sigma, final concentration of 3.5 mM), EDT (Sigma, final concentration of 1 mM) and 2-ME (Sigma, final concentration of 1 mM) and left to incubate (2 h, room temperature) followed by the addition of FlAsH-EDT_2 (Invitrogen, final concentration of 0.1 µM, 2 h, room temperature in the dark). Each sample was prepared directly into a black Corning 96 well plate which was examined in a PerkinElmer Envision 2101 multilabel plate reader (Ex filter 485 nm bandwidth 14 nm, Em filter 535 nm bandwidth 25 nm). Each pure protein was examined six times and reread three times. For the denaturant experiment, each protein was analysed as described above but with the addition of 6 M guanidine·HCl, while keeping the protein concentration the same and was incubated for 2 hours prior to the addition of TCEP, 2-ME, and EDT.

5.4.4. Medium throughput pH Vs GuHCl analysis

Each purified protein was analysed by incubating the protein (200 µl of 0.1 mg/ml) in a buffer at the relevant pH (pH 6-7, 100 mM Citrate-phosphate. pH 7.5-8.5 100 mM Tris.HCl) and with the correct concentration of GuHCl, with TCEP (Sigma, final concentration of 3.5 mM), EDT (Sigma, final concentration of 1 mM) and 2-ME (Sigma, final concentration of 1 mM) and incubated (2 h, room temperature) followed by the addition of FlAsH-EDT_2 (Invitrogen, final concentration of 0.1 µM, 2 h, room temperature). Each sample was prepared directly in a black Corning 96 well plate which was scanned in a PerkinElmer Envision 2101 multilabel plate reader (Ex filter 485 nm bandwidth 14 nm, Em filter 535 nm bandwidth 25 nm). Each protein was prepared three times separately and each plate was reread three times.

5.4.5. Medium throughput analysis of nanoparticle genesis reagents

DpsPAGCC was expressed and purified as described above and analysed by incubating the protein (200 µl of 0.1 mg/ml) in the presence of the reagent of choice. Each reagent (Chloroauric acid, Sodium borohydride, sodium Cyanoborohydride, Trisodium citrate and Ascorbic acid) was solubilised in both water, with no change to the pH, and flash buffer which was changed to pH 7.8. Varying final concentrations of each reagent was added to each well (0.01 M, 0.05 M, 0.1 M, 0.2 M, 0.5 M and 1 M) and left to incubate for an hour. To this, TCEP (Sigma, final concentration of 3.5 mM), EDT (Sigma, final concentration of 1 mM) and 2-ME (Sigma, final concentration of 1 mM) was added and incubated (2 h, room temperature) followed by the addition of FlAsH-EDT_2 (Invitrogen, final concentration of 0.1 µM, 2 h, room temperature). Each sample was prepared directly in a black Corning 96 well plate which was scanned in a PerkinElmer Envision 2101 multilabel plate reader (Ex filter 485 nm bandwidth 14 nm, Em filter 535 nm bandwidth 25 nm). Each protein was prepared three times separately and each plate was reread three times.

5.4.6. Protein characterisation

5.4.6.1. Analytical Size Exclusion Chromatography (SEC) and combined with fluorescence detection

The samples (0.5 ml of 1 mg/ml) in FlAsH buffer (100 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) were injected onto the column (GE Superdex 200 10/300 GL) at 0.5 ml/min with each protein repeated 3 times. The column was calibrated using six proteins as standards (GE Biosystems Calibration Kit).

For fluorescent monitoring, samples were prepared by adding to pure protein (0.5 ml of 1 mg/ml) in FlAsH buffer, TCEP (Sigma, final concentration of 3.5 mM), EDT (Sigma, final concentration of 1 mM) and 2-ME (Sigma, final concentration of 1 mM) and incubated (2 h, room temperature). FlAsH-EDT_2 fluorophore was added and incubated in the dark (Invitrogen, final concentration of 0.4 µM, 2 h, room temperature). This sample was desalted (GE, HiTrap 5 ml desalting column, running buffer - FlAsH buffer (100 mM Tris, 100 mM NaCl, 1 mM EDTA, pH 7.8)) before injection onto a SEC column (GE Superdex 200 10/300 GL, running buffer - FlAsH buffer (100 mM Tris, 100 mM NaCl, 1 mM EDTA, pH 7.8)). Elution volumes of 200 µl were collected during the run and placed into a black Corning 96 well plate. This plate was examined in a PerkinElmer Envision 2101 multilabel plate reader (Ex filter 485 nm bandwidth 14 nm, Em filter 535 nm bandwidth 25 nm). Each purified protein was examined three times with each plate reread three times.

5.4.6.2. Transmission Electron Microscopy (TEM)

TEM was performed on a FEI, Tecnai G^2 20, electron microscope set at 200 KeV. Proteins set to a concentration of 0.01 mg/ml in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) were immobilized on Formvar/carbon coated 3.05 mm copper grids (TAAB) and negatively stained with 1% Uranyl acetate. [12] Micrographs were analysed using ImageJ[14] (see Appendix 5.10).

5.4.6.3. Circular Dichroism Spectroscopy (CD)

The purified proteins were inspected in FlAsH buffer (100 mM Tris, 100 mM NaCl, 1 mM EDTA, pH 7.8). This experiment was performed on an Applied Photophysics LTD Chirascan spectrometer in a range of 200 nm to 260 nm with a protein concentration of 0.2 mg/ml with a path length of 0.5 mm. Thermal melts were performed on all purified proteins (0.2 mg/ml in FlAsH buffer) in a range of 4 to 85 °C [15, 16] (see Appendix 5.5, 5.6 and 5.7).

5.4.6.4. Electrospray Mass Spectrometry

Purified proteins were desalted into H_2O (GE, Hitrap 5ml desalting column) and set to a total protein concentration of 0.05 mg/ml and analysed on a Bruker MaXis mass spectrometer after the addition of 1% Formic acid (see Appendix 5.8).

5.5. References

1. Cornell, T. A.; Fu, J.; Newland, S. H.; Orner, B. P., Detection of Specific Protein–Protein Interactions in Nanocages by Engineering Bipartite FlAsH Binding Sites. *Journal of the American Chemical Society* **2013,** *135* (44), 16618-16624.
2. Grant, R. A.; Filman, D. J.; Finkel, S. E.; Kolter, R.; Hogle, J. M., The crystal structure of Dps, a ferritin homolog that binds and protects DNA. *Nature Structural Biology* **1998,** *5* (4), 294-303.
3. Zhang, Y.; Fu, J.; Chee, S. Y.; Ang, E. X. W.; Orner, B. P., Rational disruption of the oligomerization of the mini-ferritin E. coli DPS through protein-protein interface mutation. *Protein Science* **2011,** *20* (11), 1907-1917.
4. Fan, R.; Boyle, A. L.; Cheong, V. V.; Ng, S. L.; Orner, B. P., A Helix Swapping Study of Two Protein Cages. *Biochemistry* **2009,** *48* (24), 5623-5630.
5. Goodman, J. L.; Fried, D. B.; Schepartz, A., Bipartite Tetracysteine Display Requires Site Flexibility for ReAsH Coordination. *Chembiochem* **2009,** *10* (10), 1644-1647.
6. Scheck, R. A.; Lowder, M. A.; Appelbaum, J. S.; Schepartz, A., Bipartite Tetracysteine Display Reveals Allosteric Control of Ligand-Specific EGFR Activation. *Acs Chemical Biology* **2012,** *7* (8), 1367-1376.
7. Avraham, R.; Yarden, Y., Feedback regulation of EGFR signalling: decision making by early and delayed loops. *Nature Reviews Molecular Cell Biology* **2011,** *12* (2), 104-117.
8. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **2004,** *25* (13), 1605-1612.
9. Scheck, R. A.; Schepartz, A., Surveying Protein Structure and Function Using Bis-Arsenical Small Molecules. *Accounts of Chemical Research* **2011,** *44* (9), 654-665.
10. Inc, A. S., Discovery Studio Modeling Environment, Release 4.0, San Diego. **2013**.
11. Luedtke, N. W.; Dexter, R. J.; Fried, D. B.; Schepartz, A., Surveying polypeptide and protein domain conformation and association with FlAsH and ReAsH. *Nature Chemical Biology* **2007,** *3* (12), 779-784.
12. Fan, R.; Chew, S. W.; Cheong, V. V.; Orner, B. P., Fabrication of Gold Nanoparticles Inside Unmodified Horse Spleen Apoferritin. *Small* **2010,** *6* (14), 1483-1487.
13. Zhang, Y.; Raudah, S.; Teo, H.; Teo, G. W. S.; Fan, R.; Sun, X.; Orner, B. P., Alanine-shaving Mutagenesis to Determine Key Interfacial Residues Governing the Assembly of a Nano-cage Maxi-ferritin. *Journal of Biological Chemistry* **2010,** *285* (16), 12078-12086.
14. Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W., NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **2012,** *9* (7), 671-675.
15. Kelly, S. M.; Jess, T. J.; Price, N. C., How to study proteins by circular dichroism. *Biochimica Et Biophysica Acta-Proteins and Proteomics* **2005,** *1751* (2), 119-139.
16. Greenfield, N. J., Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols* **2006,** *1* (6), 2876-2890.

Chapter 6

Bipartite tetracysteine display for high throughput screening of Dps libraries in living bacteria

6.1. Introduction

The development of a Dps protein nanocage self-assembly detector, which was used to assay quaternary stability in conditions needed for gold nanoparticle formation in a medium throughput manner (Chapter 5), was a significant step forward in nanocage research. This technique was used to quickly allow for the detection of nanocages without the need for protein purification. The potential applications of such a technology could be for nanoparticle formation as previously mentioned but also to speed up research to engineer protein interfaces and map key or 'hot spot' residues. The major limitation of this technique is that it is currently limited to *in vitro*, limiting the number of proteins or conditions that can be examined at any one time.

This assay could be used for other techniques rather than just for characterisation after a mutant design has been cloned and expressed. In fact, such a technology which can detect one state or species over another could be adapted to screen and separate a mixture of mutants based on which state they belong to. By examining the alterations that could push proteins from one state to another, data can be gathered about the structural attributes or residues that are needed for protein-protein interactions, helping to resolve the reasons for quaternary structure formation. Not only could fundamental data be obtained unravelling how proteins interact, but potentially the ability to separate different states, could allow for the development of new and novel engineered traits, such as newly engineered interactions or functions. However, to successfully gather enough data or screen enough engineered conditions; any approach must be taken into a high throughput manner. The easiest way to perform this is to use cells that contain overexpressed proteins which bind to or exhibit the property that allows for their separation. This way, with a 'one cell/one clone' format, entire cells can be separated and their contents analysed to reveal the genetic differences that have allowed for their selection and separation. This can then allow for the collection of mass data pertaining to quaternary structure.

6.1.1. Aims of this investigation

While a method for the direct detection of the self-assembly of the Dps nanocage was reported in Chapter 5, this method was limited to *in vitro* conditions and although this assay worked in cellular like lysate environments, the throughput of this approach was still limited. The aim of this investigation is to move that assay into *in vivo*, for use in bacterial cells with over expressed proteins. Methods such as microscopy and flow cytometry were employed to see the extent of fluorescence in cells, but also the efficiency of the labelling process. A further goal of this investigation was to see if this assay could be used in a high throughput manner to screen libraries of mutants to see if specific nanocage characteristics could be selected over others.

6.1.2. Dps as a model system for screening libraries for protein cage stability

It was decided that Dps would act as the proof of concept system for the expansion of our technique to the screening of protein libraries for nanocage stability in living cells. Chapters 4 and 5 described our efforts in developing the method *in vitro* for Bfr and Dps, respectively, with the latter protein exhibiting behaviour much more amenable to our conditions. In addition, extensive research in our laboratory on Dps provides insight and controls for the development of the technique *in vivo*. For example, we used alanine scanning mutagenesis to identity 'hot spot' residues along the interfaces of Dps which provided us with Dps mutants that exhibit a range of nanocage to dimer intermediate ratios in contrast to wild-type which forms nanocages predominantly [1,2]. One mutant that is used chapter, is a Dps double mutant (R83A and R133A) that assembles cleanly into 100% dimers. These arginine's are located near the 'Dps like' three-fold axis of symmetry at the interface between two monomers and form an extended hydrogen network with three aspartic acid residues. Disruption of this network results in disassembly of the nanocage state (Figure 6.1). Thus, along with the wild-type protein as a control for 100% presentation of nanocage, this project used double mutants at positions 83 and 133 as dimer-only controls. The double alanine, as

described above, along with double aspartate and double tryptophan, which should inhibit nanocage formation by coulombic and steric repulsion respectively, were all developed in this regard (see sections 6.2.1.1 and 6.2.1.3.).

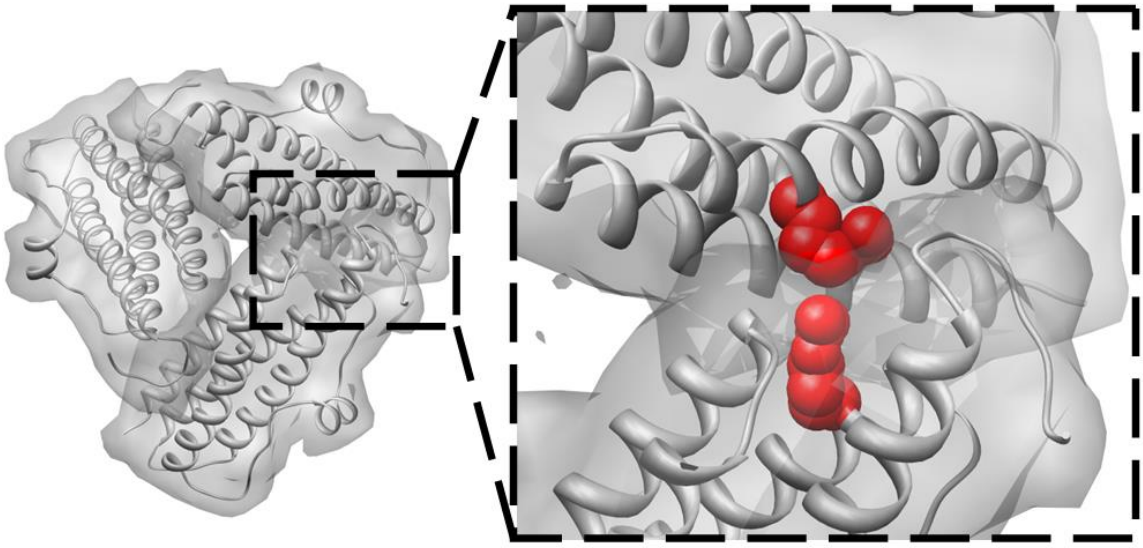


Figure 6.1 – Protein crystal structure for Dps (PDB:1DPS) showing the location of the mutants R83 and R133 (red) at a monomer-monomer protein-protein interface near the Dps-like three-fold axis of symmetry. Double mutation of these residues to alanine results in Dps that cleanly forms dimer. Proteins with double mutations at these positions will be the basis of library screens described in this chapter. Figure made using Chimera software[3].

6.1.3. FlAsH technology with living cells

Expansion of our method *in vivo* would likely require re-optimisation of conditions from those discussed in Chapters 4 and 5[4] therefore the details of the procedure were reassessed. Since its early development, FlAsH-EDT_2 has been used for monitoring proteins inside cells, and multiple conditions have been employed[5, 6, 7]. As discussed in Chapter 1, most cellular work performed with FlAsH-EDT_2 was performed in mammalian cells. This can provide us with useful information into the requirements and limitations of FlAsH, however, it is intended that *E. coli* is used throughout this investigation. Only a small body of work has involved the use of FlAsH technology in bacterial cells. Most notably, Van Nhieu and co-workers labelled proteins expressed in bacteria with FlAsH, and they characterised the effect

of FlAsH on the viability of the cell[8]. This, along with other work[9, 10, 11] described in Chapter 1, provides a foundation from which to expand the Dps nanocage detection assay into living bacterial cells.

6.1.4. Generation of protein nanocage libraries

The design, generation and characterisation of individual mutants is slow and labour-intensive, thus limiting the number of designs that can be explored. The development of *in vivo* conditions, which is an aim of the research described in this chapter permits direct detection of the oligomerization state of the protein inside of the cell bypassing the necessity of purification. The second aim of this research is to develop a high throughput screen of protein libraries to discover proteins with enhanced nanocage forming ability. Fluorescently activated cell sorting (FACS) allows for the fluorescent analysis of individual cells followed by their collection. In principle, this would provide us with a method to screen libraries. We could express single protein clones of a library in single bacteria, label these clones with FlAsH, and then use FACS to identify and isolate individual living bacteria that are most fluorescent and presumably are expressing the most nanocage-prone library member. Because we would be screening living cells, we would also be able to amplify after screening and run the evolved library back through additional rounds of screening. We could also imagine performing mutagenesis between screening rounds to enhance diversity. Because each clone is expressed inside a bacterium which encodes its sequence, a simple "identification tag" is intrinsically provided, and simple DNA sequencing could be used to "deconvolute" hits. It could be imagined that high-content sequencing strategies could be used to follow the evolution of the library over multiple rounds in the future.

The "one clone/one bacterium" (OCOB) FACS screening technique is flexible enough to be coupled to various methods for diversity generation. As described in Chapter 1, several cloning methods have been developed which introduce sites of mutational diversity either all

along the length of a gene[12], the entire plasmid[13], or at selective sites[14]. These methods can be used to create diverse libraries of designs often in one pot molecular biology reactions. One method, transfer polymerase chain reaction[15] (TPCR, see Chapter 1), is harnessed in this chapter to generate a library of Dps mutants. The TPCR technique allows us to diversify specific residues in key protein-protein interactions semi-rationally. This method was chosen because it introduces mutations selectively to positions separated along the primary sequence, and cloning success can be easily followed by agarose gel.

The strategy pursued in this chapter will be to first expand our protein nanocage stability fluorescence assay into living bacteria by further optimising conditions using well-characterised control proteins. When it is characterised, semi-rationally designed Dps mutant libraries will be generated and screened to identify proteins that assemble into more stable nanocages. It is thought that these resultant mutants could form a foundation of bypassing problems experienced in Chapter 2 along with a multitude of other applications, possibly some currently unforeseen, emphasizing the enabling aspect of this technology.

6.2. Results and discussions

6.2.1. Expansion of Dps oligomerization assay to *in vivo* conditions

6.2.1.1 Generation of single constructs in pET-46 Ek/LIC to provide nanocage forming proteins both *in vivo* and that can also be easily purified

Throughout Chapters 4 and 5, two different plasmids were used. One plasmid, pET-22b, was employed for expression of the target proteins without any fusions to enhance purification or stability and the second, pET-32b, provided expressed proteins fused to three tags at the N-terminus (Trx-tag, S-tag and His_6). While the tags of pET-32b are necessary to purify the proteins, these large fusions can interfere with nanocage assembly. Thus, this plasmid also provides an Enteriokinase protease (Ek) site to remove the tags after purification. While this does not hinder any *in vitro* assays that take place post digestion, it hinders experiments in lysate conditions or *in vivo* with this plasmid and necessitated pET-22b for the lysate experiments. This strategy, while effective, was inefficient. It was thought that a method allowing us to directly express, isolate, and characterise *in vitro* a "hit" clone after performing the FlAsH/FACS/OCOB screening without requiring re-cloning would be more ideal.

To do this, we opted to express the proteins from a single plasmid, pET-46 Ek/LIC, that provides only a short, Ek-cleavable, His_6 purification tag at the N-terminus, which, due to its small size should not interfere with nanocage assembly. This strategy would permit *in vivo* selection followed by *in vitro* characterisation without the time consuming re-cloning process. It should be noted that this approach may prove non-general to other protein systems because a) the N-terminus of Dps is projected outside the nanocage and thus a small fusion at this position is not expected to greatly affect a protein-protein interaction b) the designed FlAsH binding site established in Chapter 5 is located at the Dps C-terminus.

Thus, the genes for all designs and controls described in this Chapter were cloned into pET-46 Ek/LIC (Figure 6.2). These included the positive control, **DpsCCPGCC**, negative control, **Dps**, and most promising design, **DpsPAGCC**, from Chapter 5. In addition, expected dimer forming mutants, **Dps(AA)PAGCC**, **Dps(DD)PAGCC** and **Dps(WW)PAGCC**, (see section 6.2.1.3. for details) were cloned as additional controls. The TPCR library was generated on the **Dps(AA)PAGCC** pET-46 Ek/LIC control as a background sequence. Screening of the library resulted in the mutants **Dps(AA)PAGCC**^{Q86V,L87F,N130F}, **Dps(AA)PAGCC**^{N130Y} and **Dps(AA)PAGCC**^{N130L,D141D,D142W,D143G,} that were isolated and characterised *in vitro* (see section 6.2.2.3. for details) (Figure 6.2). It should be noted that when the proteins were isolated for *in vitro* analysis, the His_6 purification tag was not removed so as to result in more rapid and higher yielding purification. Therefore to ensure these slightly altered variants still assembled correctly and bound to FlAsH as previously determined, the control proteins were expressed, purified and fully characterised (see below and Appendices 6.2 and 6.3 for SDS-PAGE and Mass spectrometry respectively).

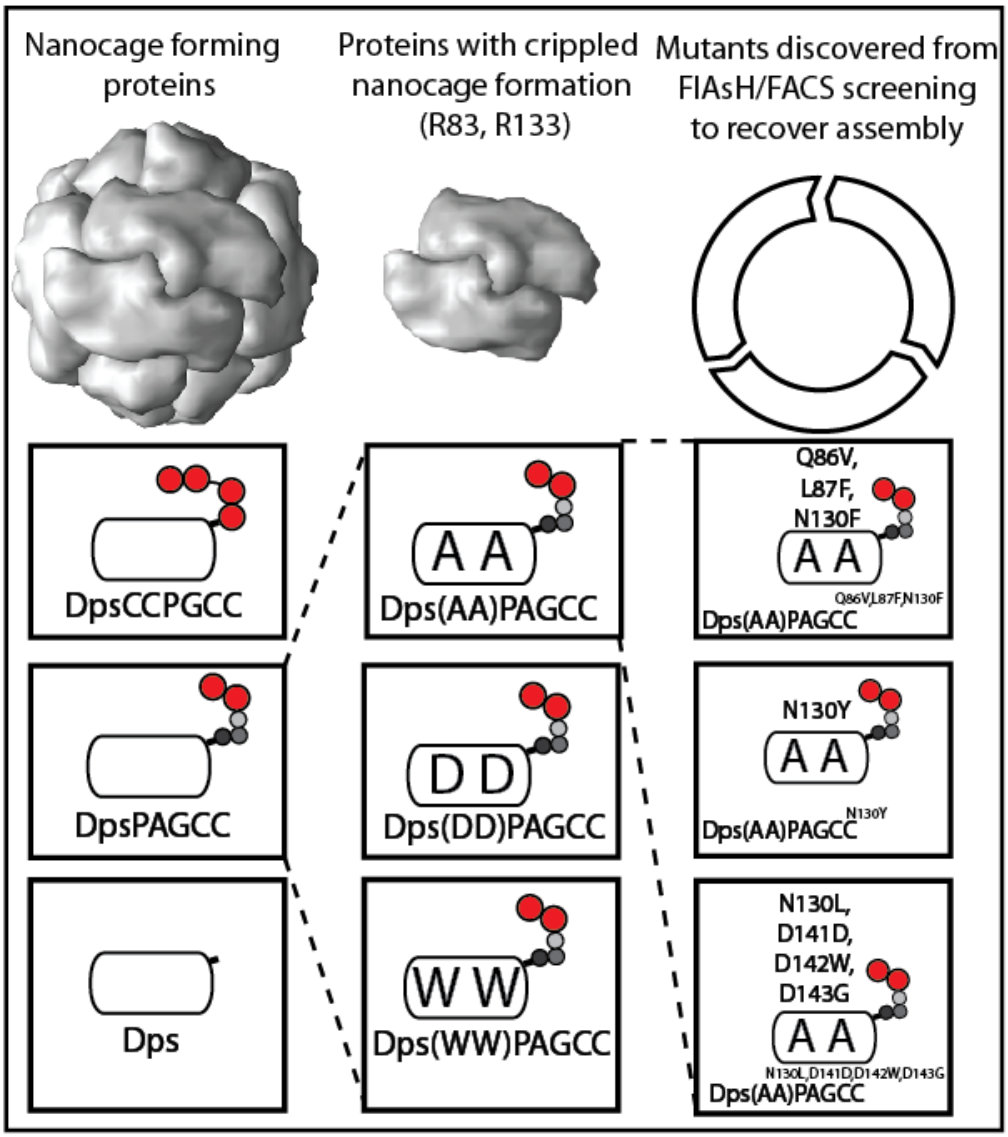


Figure 6.2 – Proteins used in the research described in this chapter. (left column) Dps crystal structure (PDB: 1DPS) and all nanocage forming designs and controls as described and characterised in Chapter 5: the positive control (**DpsCCPGCC**) with full tetracysteine FlAsH binding site, the most promising bipartite design (**DpsPAGCC**) and the negative control with no additional cysteines (**Dps**). (middle column) Control proteins designed to have crippled assembly properties (See section 6.2.1.3 for details). Each contains mutations at R83 and R133, including double alanine, **Dps(AA)PAGCC**, double aspartic acid, **Dps(DD)PAGCC**, and double tryptophan, **Dps(WW)PAGCC**. (right column) Mutants with recovered assembly properties obtained from FlAsH/FACS screening of an OCOB library based on **Dps(AA)PAGCC** (See section 6.2.2.3. for details). Cysteine (red), glycine (light grey), alanine (dark grey), proline (black). For crippled assembly controls; alanine (A), aspartic acid (D), tryptophan (W).

The three proteins characterised in Chapter 5 (**DpsCCPGCC**, **DpsPAGCC**, **Dps)** showed nanocage like architectures in both SEC (Figure 6.3) and TEM (See Appendix 6.4) and the purified proteins demonstrated fluorescence intensities with FlAsH similar to those previously observed (Figure 6.4) (see 6.2.1.3. for discussion of dimer control analysis with **Dps(AA)PAGCC**,

Dps(DD)PAGCC, and **Dps(WW)PAGCC** and 6.2.2.3 for discussion of library design, sorting and characterisation).

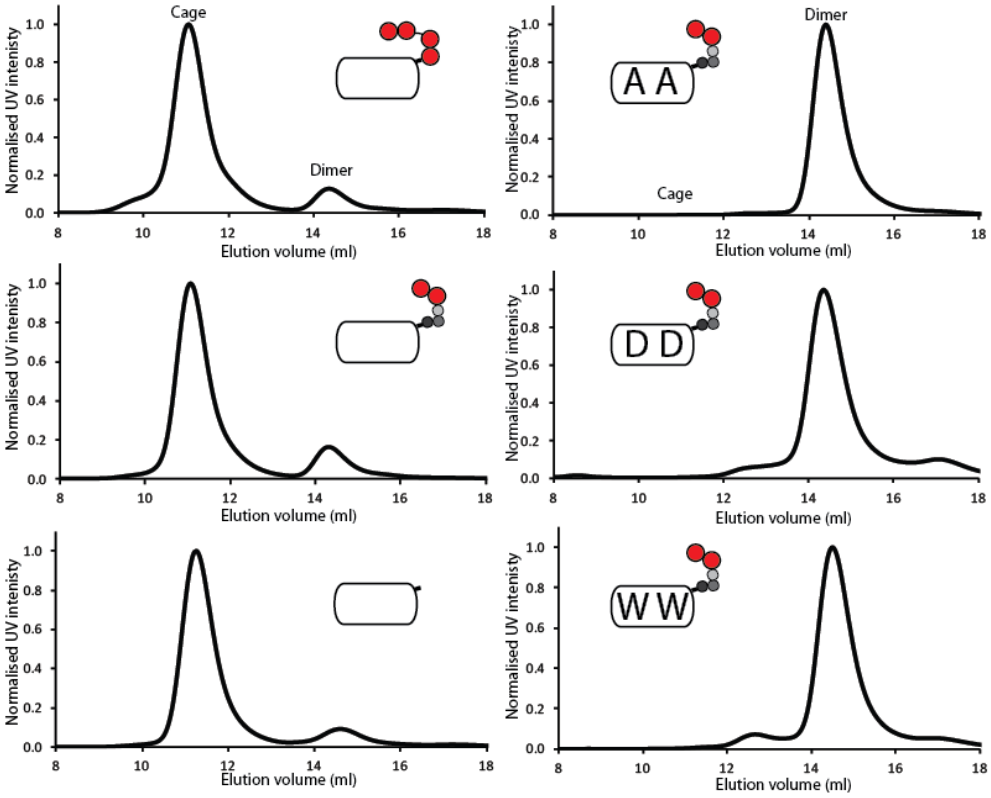


Figure 6.3 – Size exclusion chromatograms of purified Dps derivatives expressed in pET-46 Ek/LIC. (top left) positive control **DpsCCPGCC**, (middle left) most promising bipartite design **DpsPAGCC**, (bottom left) negative control **Dps**. (top right) **DpsPAGCC** with cage-crippling mutations R83A and R133A (**Dps(AA)PAGCC**), (middle right) **DpsPAGCC** with cage-crippling mutations R83D and R133D (**Dps(DD)PAGCC**), (bottom right) **DpsPAGCC** design with cage-crippling mutations R83W and R133W (**Dps(WW)PAGCC**). Traces are averages of three runs, normalised to the highest intensities for each (0.5 ml injection of 0.5 mg/ml protein). For explanation of cage-crippled mutants see Section 6.2.1.3. See Figure 6.2 for explanation of schematics.

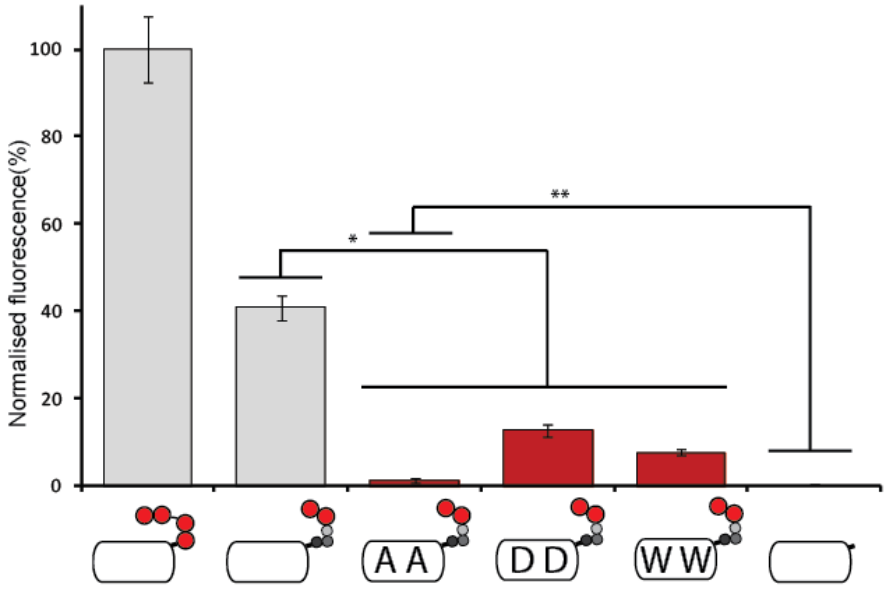


Figure 6.4 - FlAsH fluorescence of purified Dps derivatives expressed in pET-46 Ek/LIC. The data is normalised to the positive and negative controls. All proteins are 0.1 mg/ml in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) with 0.1 µM FlAsH, 3.5 mM TCEP, 1 mM EDT, 1 mM 2-ME. Data is averaged from 8 replicates. Error bars are S.D* Two-tailed P-values = 0.0001. ** Two-tailed P-values = 0.0023. See Figure 6.2 for explanation of schematics.

These data verified that expressing these proteins with the small His_6 tag from the new plasmid had little impact on either nanocage assembly or FlAsH binding *in vitro*. The next step, therefore, was to analyse the ability of FlAsH to bind selectivity to these control proteins in living *E. coli*.

6.2.1.2. Dps nanocage assembly evaluated *in vivo* with fluorescence microscopy.

To gain an initial indication of whether or not FlAsH-EDT_2 can be used to selectivity label our designed proteins in living bacterial cells, individual bacterium were assessed with fluorescence microscopy. Nhieu and co-workers[9] analysed the toxicity of FlAsH-EDT_2 on gram negative bacteria and found little effect in concentration ranges up to 20 µM. They were also able to FlAsH label intracellular proteins tagged with CCPGCC by incubating the cells in a solution containing high concentrations of reducing agent (TCEP)[8]. However, Gierasch and co-workers showed that intracellular proteins in *E. coli* could also be labelled by first treating the cells with lysozyme to permeabilize the bacteria. With both of these conditions in mind, we developed new conditions which included high concentrations of FlAsH-EDT_2 (20 µM) and the

dithiol EDT (2 mM). As the cytosol of the bacterium is already in a reduced condition, it was deemed unnecessary to add either of the reducing agents seen in Chapter 5 (TCEP or 2-ME). The Dps controls (**DpsCCPGCC** and **Dps**) were exposed to these conditions and analysed with fluorescence microscopy (Figure 6.5). As expected micrographs of bacteria expressing **DpsCCPGCC** showed a strong fluorescence response whereas control cells expressing **Dps** exhibited almost no detectable fluorescence in identical conditions. Taken together these data suggest that there is low background signal, that FlAsH-EDT_2 can successfully penetrate these cells and FlAsH fluorescence is specific for the tetracysteine tag. It also solidly establishes our controls.

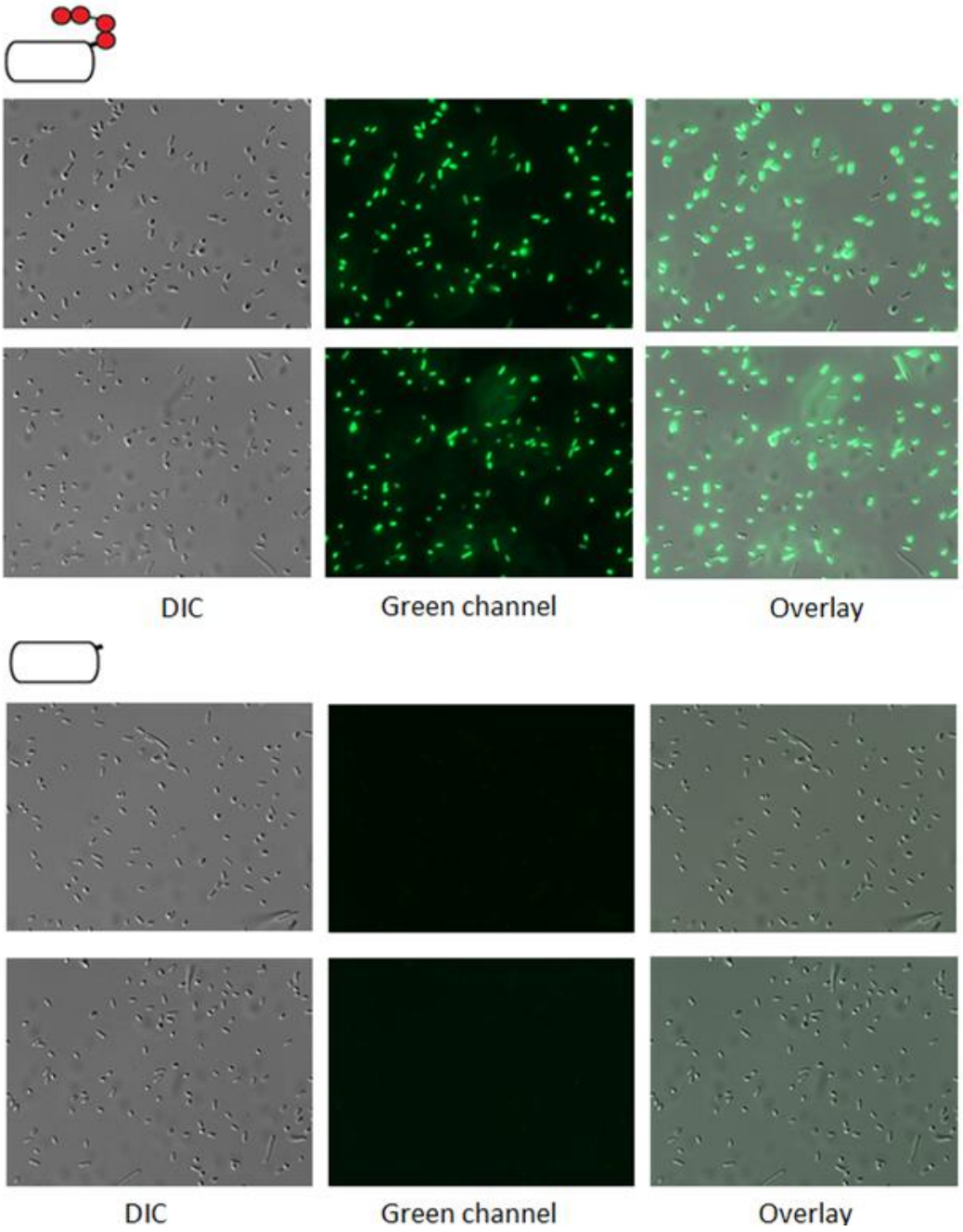


Figure 6.5 – Analysis of intracellular FlAsH labelling of overexpressed Dps protein controls in *E.coli* with epi-fluorescence microscopy. (top) Positive control cells expressing **DpsCCPGCC**, (bottom) negative control cells expressing **Dps** (left to right) DIC image (100x magnification, 20 ms exposure), green channel (50 ms exposure, filter sets Ex – 490 ±20 nm, Em – 525 ±36 nm), overlaid image. Two micrographs of each are included to demonstrate reproducibility. After 1.5 hours of protein expression, the cells were treated with 20 µM FlAsH-EDT_2 in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH7.8) with 2 mM EDT. Images generated with ImageJ[16]. Gain was not changed between images of the controls.

With controls firmly validated, cells expressing **DpsPAGCC** was subjected to identical microscopic analysis (Figure 6.6). These cells were clearly fluorescent however with intensities that were less pronounced than those from cells expressing **DpsCCPGCC**. This result

demonstrates that bipartite tetracysteine display can be used in bacterial cells for the detection of nanocage self-assembly.

It should be noted, however, that several of the cells in the micrographs of both the **DpsCCPGCC** and **DpsPAGCC** expressing cells had very little associated fluorescence. There could be a number of explanations for this result, including selective protein expression, contamination, cell death, or differentiation. Although this may suggest a limitation with microscopy for the determination of how many of the cells are labelled and to what extent, other methods, such as flow cytometry could complement this technique by providing more quantitative data.

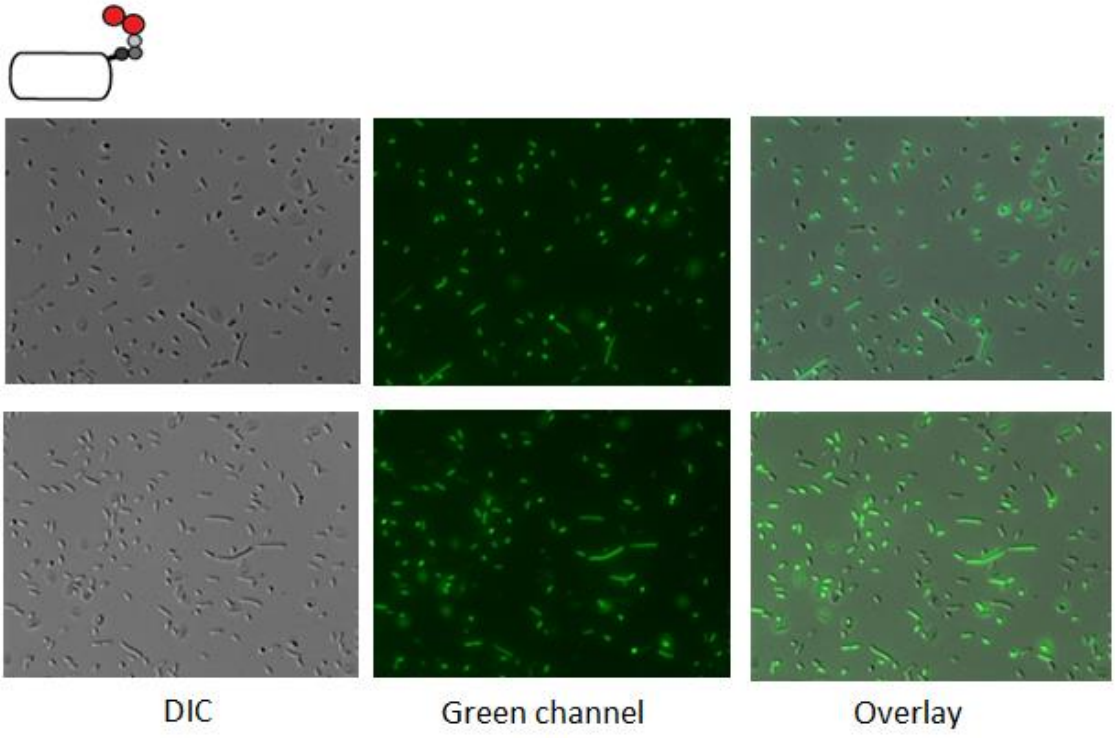


Figure 6.6 - Analysis of intracellular FlAsH labelling of overexpressed **DpsPAGCC** in *E.coli* with epi-fluorescence microscopy. (left to right) DIC image (100x magnification, 20 ms exposure), green channel (50 ms exposure, filter sets Ex – 490 ±20 nm, Em – 525 ±36 nm), overlaid image. Two micrographs of each are included to demonstrate reproducibility. After 1.5 hours of protein expression, the cells were treated with 20 µM FlAsH-EDT_2 in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH7.8) with 2 mM EDT. Images generated with ImageJ [16]. These images were collected in identical conditions, on the same day and with the same microscope as those in Figure 6.5, with no change in gain.

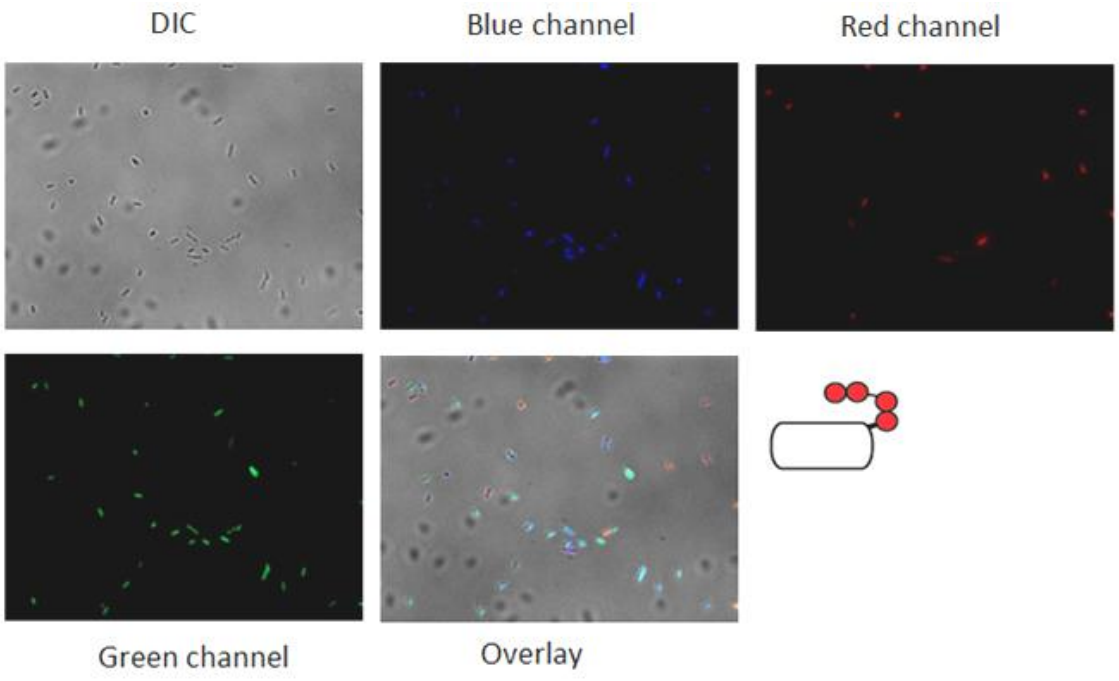
Before flow cytometry was attempted, and to provide additional controls for the cytometry, the labelling method was further examined with microscopy to determine the

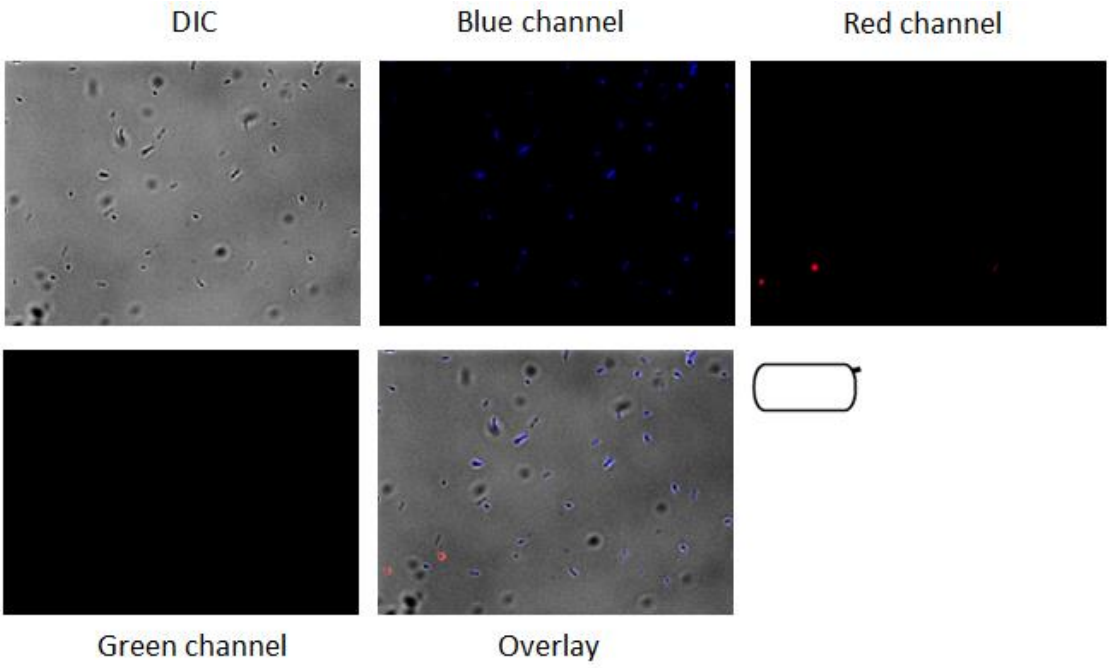
impact it has on bacteria viability because the high throughput screening of OCOB libraries with FACS requires that the bacteria survive the process. To investigate this, a bacterial dead/live screen was envisioned using two additional fluorophores. Hoechst 33342, which fluoresces upon DNA binding, was used to identify all (living and dead) bacterial cells and propidium iodide (PI), which can only penetrate cells with disrupted membranes, identified dead cells. While this is not precisely a dead/live screen, this method can identify all cells, which is a useful tool when removing background signals from flow cytometry, and separately identify dead cells which can then be removed leaving only the live cells (Hoescht^{+}, PI^{-}). In addition the number of PI-positive cells would indicate the extent to which the cells are being killed. One consideration for the selection of these probes, is spectral overlap (Table 6.1). This combination of fluorophores should allow for a three channel excitation system using the blue, green and red channels with little overlap.

Fluorophore	Excitation maxima (nm)	Emission maxima (nm)	Channel	Detection role
Hoechst 33342	350	461	Blue	Bacterial cells
FlAsH-EDT$_2$	508	528	Green	Nanocages
Propidium Iodide	535	617	Red	Dead cells

Table 6.1– Excitation and emission maxima for all fluorophores used in the research described by this chapter (Invitrogen).

To confirm that each fluorophore can be observed in our conditions with these cells, the microscopy method described above was repeated with all three probes (Figure 6.7) and the three proteins. The response from the green channel showed a similar trend to that seen above when just using a single fluorophore. A response was seen in both the blue and red channels, with most of the cells being Hoechst positive and only a limited number of cells were PI positive, meaning that the majority of cells survived this process and were alive. Moreover, the presence of a response from all three probes showed that this system could work to not only identify whether a nanocage is formed but also whether or not the cell is alive. With this method characterised in cells using microscopy, it was important to analyse the level of labelling quantitatively with flow cytometry.


DIC
Blue channel
Red channel
Green channel
Overlay


DIC
Blue channel
Red channel
Green channel
Overlay

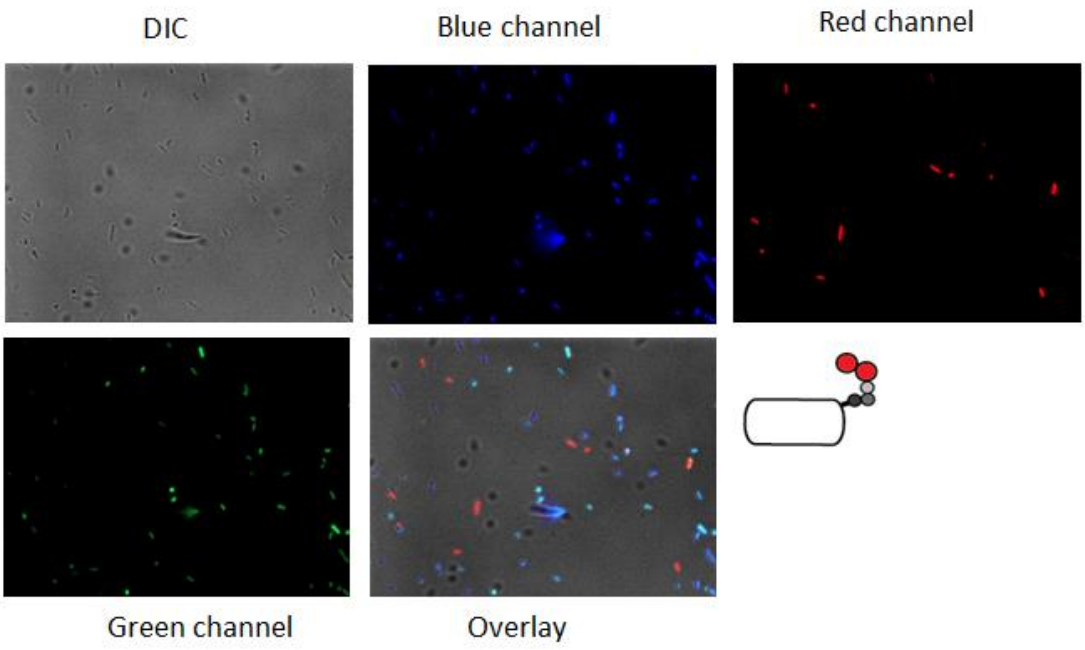


Figure 6.7 - Analysis of intracellular labelling of the three fluorophore system for bacteria expressing Dps protein controls and designs with epi-fluorescence microscopy. (top) cells expressing positive control **DpsCCPGCC**, (middle) negative control **Dps,** (bottom) bipartite design **DpsPAGCC**. (left to right of top row) DIC image (100x magnification, 20 ms exposure), blue channel (2 s exposure, filter sets Ex – 402 ± 15 nm, Em – 455 ± 50 nm), red channel (20 ms exposure, filter sets Ex – 555 ± 25 nm, Em – 605 ± 52 nm). (left to right bottom row) green channel (50 ms exposure, filter sets Ex – 490 ± 20 nm, Em – 525 ± 36 nm) and overlaid image. After 1.5 hours of protein expression, the cells were treated with 20 µM FlAsH-EDT_2 in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH7.8) with 2 mM EDT. After a further 1.5 hours, 0.1 µM Hoechst 33342 and 0.03 µM PI were added. Images generated with ImageJ [16].

6.2.1.3. Dps nanocage assembly evaluated *in vivo* with flow cytometry

Flow cytometry measures the fluorescence intensity of each cell in multiple channels. While the Hoechst and PI fluorophores could be used to select cells that are alive during this investigation, and aspect which is important if these cells were to be sorted and collected, the FlAsH-EDT_2 probe assesses the ability of any proteins to self-assemble into nanocages. First, however, it was necessary to confirm that the cells can be selected and that dead ones can be screened out. This was accomplished by applying the two fluorophore, dead/alive screen that was described above to flow cytometry (Figure 6.8).

To determine the ability of cytometry to identify the live and dead populations, cells expressing the positive control protein **DpsCCPGCC**, was treated with either PI, Hoechst 33342, or both. (Figure 6.8) The fluorescence intensity in the blue channel increases dramatically for

the entire population when Hoechst 33342 is added indicating that this dye is able to identify all cell populations, however, only a small shoulder appears in the red channel when PI is administered, indicating a small population of dead cells. This population becomes the majority when the cells are killed by heat treatment as would be expected confirming that the shoulder in the red channel of un-heat treated cells is, in fact, due to dead cells. The fact that the population of dead cells when they are not heat treated is small, indicates that the conditions we are using for FlAsH-EDT_2 labelling are non-lethal.

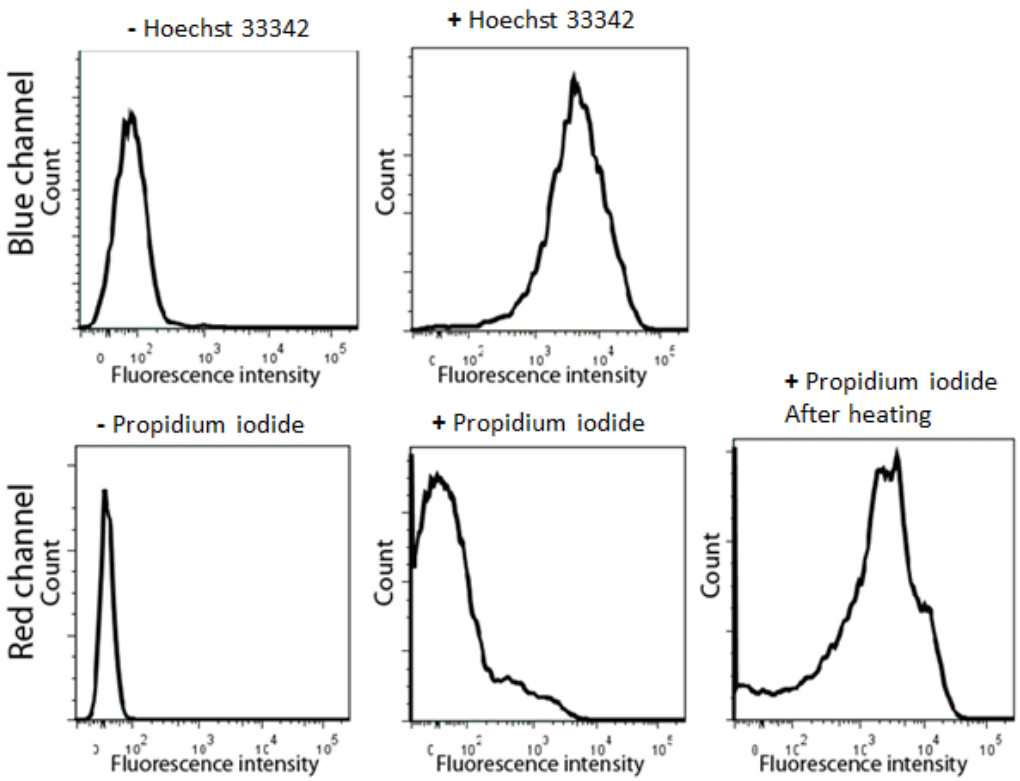


Figure 6.8 – Dead/live analysis using flow cytometry with cells expressing the Dps positive control **DpsCCPGCC**. (left top and bottom) Fluorescence in red and blue channels of population of bacteria expressing **DpsCCPGCC** with no added PI or Hoechst 33342. (middle top) Fluorescence in blue channel in the presence of Hoechst 33342. (middle bottom) Fluorescence in the red channel in the presence of PI, (right bottom) Fluorescence in the red channel in the presence of PI after cells were killed by heating at 60 ^{0}C for 5 minutes. After 1.5 hours of protein expression, samples changed into FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH7.8) with 2 mM EDT. After a further 1.5 hours, 0.1 µM Hoechst 33342 and/or 0.03 µM PI were added if indicated. Blue channel: Ex laser - 375 nm, Em filter- 450 ±20 nm. Red channel: Ex laser – 488 nm, Em filter – 670 ±20 nm. Data normalised through software (FlowJo, Tree star) using 5000 cells.

Before discussing how this was applied to assessing nanocage formation, a note about how the data will be handled is appropriate. (Figure 6.9) Bacteria probed with all three fluorophores will first be analysed in the blue channel and Hoechst-positive events will be gated before analysis in the red channel. Of these, PI-positive (i.e. dead cells) will be noted and PI-negative events will be gated out for analysis in the green channel. The Hoechst-positive, PI-negative, FlAsH-positive events would be gated out and collected if performing cell sorting (Figure 6.9).

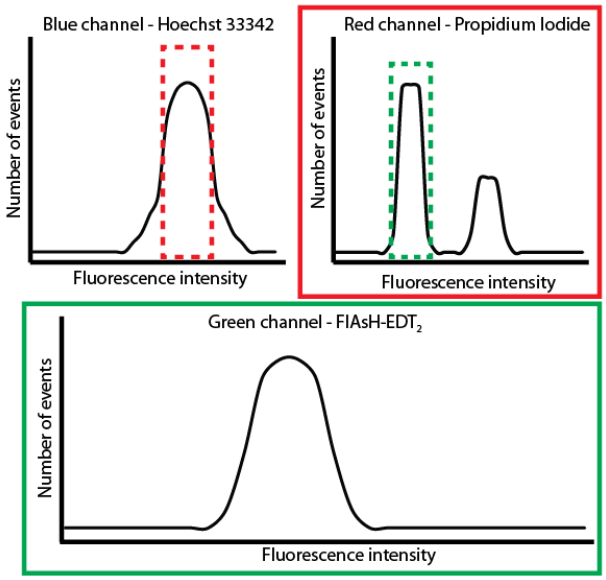


Figure 6.9 – Schematic representation of data treatment in flow cytometry analysis of bacteria probed with Hoechst 33342, PI and FlAsH. (top left) First the blue channel is analysed and the Hoechst$^+$ cells gated (red dashed box). This population is analysed in the red channel (top right) and the live, PI$^-$, cells (green dashed box) are gated away from the PI$^+$, dead cells. (bottom) The live population is then further interrogated in the green channel for binding to FlAsH to identify the FlAsH$^+$ cells that express proteins that can self-assemble into nanocages.

The flow cytometry analysis using the method described above demonstrated that baseline separation was evident between the positive, **DpsCCPGCC,** and negative controls, **Dps,** (Figure 6.10a) indicating a sufficient window to identify clones expressing proteins of varying nanocage stabilities. Satisfyingly consistent with the microscopy and lysate data (above and Chapter 5), the cells expressing **DpsPAGCC** displayed a distribution that was intermediate between the controls demonstrating that it is possible to use flow cytometry in living cells to

identify proteins that assemble into nanocages (Figure 6.10b). While this data is extremely promising, if this strategy is to be used to screen OCOB libraries, it needs to be able to distinguish clones that express bipartite designs that have different levels of stability. To test this, we used Dps mutants that are crippled in their ability to form the nanocage oligomerization state.

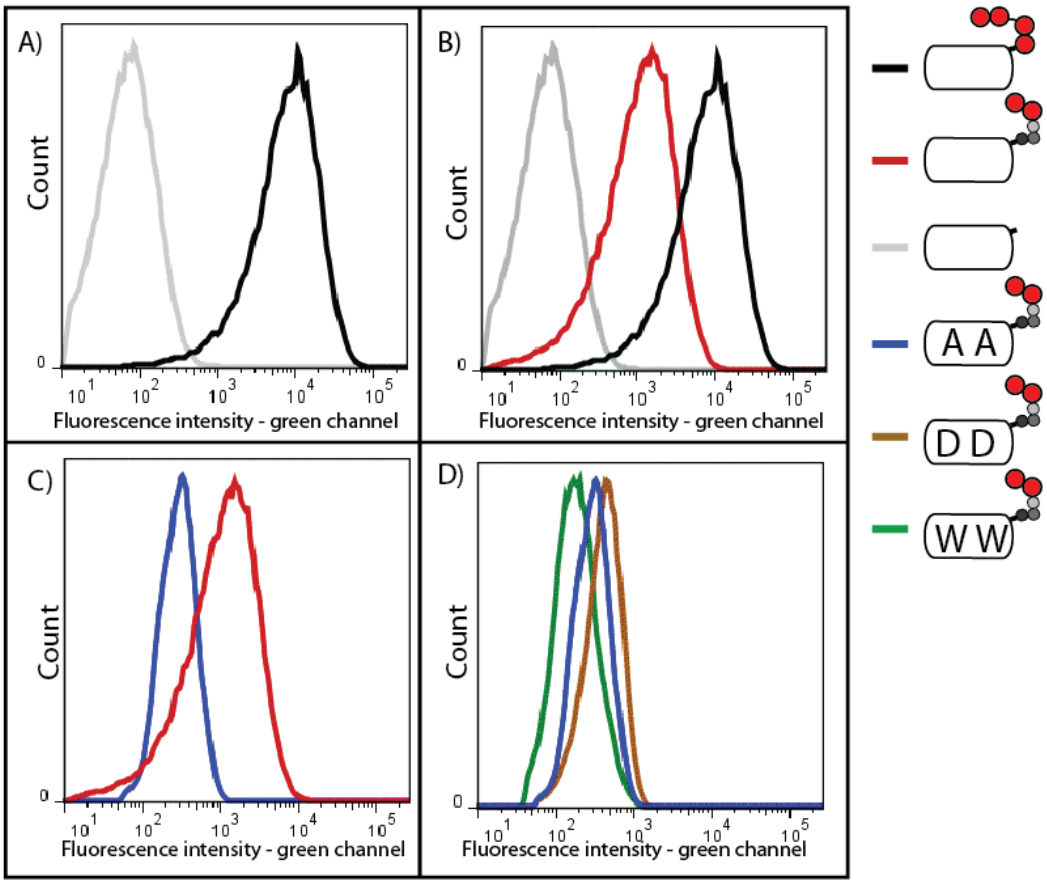


Figure 6.10 – Flow cytometry to assess the oligomerization state of Dps derivatives. (A) Fluorescent events for populations of cells expressing negative control, **Dps** (light grey), and positive control, **DpsCCPGCC** (black), (B) comparison of clones expressing Dps controls to those expressing **DpsPAGCC** (red). (C) Comparison of clone expressing **DpsPAGCC** (red) to to one that expresses the cage crippled control, **Dps(AA)PAGCC** (blue). D) Comparison of cage crippled controls **Dps(AA)PAGCC** (blue), **Dps(DD)PAGCC** (brown) and **Dps(WW)PAGCC** (green). After 1.5 hours of protein expression, the cells were treated with 20 µM FlAsH-EDT_2 in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH7.8) with 2 mM EDT. After a further 1.5 hours, 0.1 µM Hoechst 33342 and 0.03 µM PI were added if indicated. Blue channel: Ex laser - 375 nm, Em filter- 450 ±20 nm. Red channel: Ex laser – 488 nm, Em filter – 670 ±20 nm. Data normalised through software (FlowJo, Tree star) using 100,000 cells.

Because an early goal for developing a FlAsH/FACS OCOB high throughput screen would be to discover proteins with enhanced nanocage forming ability, it was necessary to

determine if it was possible to distinguish proteins with different nanocage stabilities and to optimise the relative difference between their oligomerization-dependent signals. Previously, our laboratory performed alanine shaving mutagenesis on the wild-type protein Dps to identify residues important for nanocage assembly[17]. Several positions were found that lowered the melting temperature of Dps as well as alter the ratio between nanocage and dimer. A double mutant, Dps(R83A, R133A) involving residues located near the Dps-like three-fold axis of symmetry (Figure 6.1), formed 100% dimer. We thought this double mutant could be the basis of the next step of development. These mutations were placed onto the bipartite Dps design resulting in **Dps(AA)PAGCC** which was expected to assemble into 100% dimer. This protein was cloned, expressed, purified, and characterised (see Appendices 6.2 and 6.3 for SDS-PAGE and mass spectrometry respectively). As expected, this protein showed no nanocage-like structures in either SEC (Figure 6.3) or during the forcing conditions of TEM (see Appendix 6.4). In addition, it exhibited little fluorescence signal in the *in vitro* FlAsH assay (Figure 6.4) compared to **DpsPAGCC**. **Dps(AA)PAGCC** was further characterised with flow cytometry (Figure 6.10c). Cells expressing **Dps(AA)PAGCC** and treated with FlAsH exhibited an average fluorescence intensity lower than that of **DpsPAGCC** but higher than **Dps**. This decrease in signal compared to **DpsPAGCC** was promising; however, the fact that it was higher than the negative control, **Dps,** was somewhat unexpected because very little nanocage state was observed *in vitro*. It is possible that this signal is caused by a population of nanocages inside the bacteria where high concentrations of proteins are "pushing" the nanocage together from dimer intermediates. Alternatively, protein aggregates could be responsible for the higher than expected signal, however, very little aggregation was observed in the SEC and during purification. Therefore, further investigation of **Dps(AA)PAGCC** was undertaken in lysates with high total protein concentrations to try to replicate cellular conditions.

To determine if the higher than expected signal for **Dps(AA)PAGCC** *in vivo* was due to an increase in protein concentration, different total lysate protein concentrations were

scanned (Figure 6.11). While a significant change in signal is observed at different protein levels, it is not concentration dependent. The lack of a clear trend highlights some of the difficulties when working with lysate samples especially at high protein concentrations. Another approach to explore this issue was to create mutants that are even more crippled in their ability to form the nanocage oligomerization state.

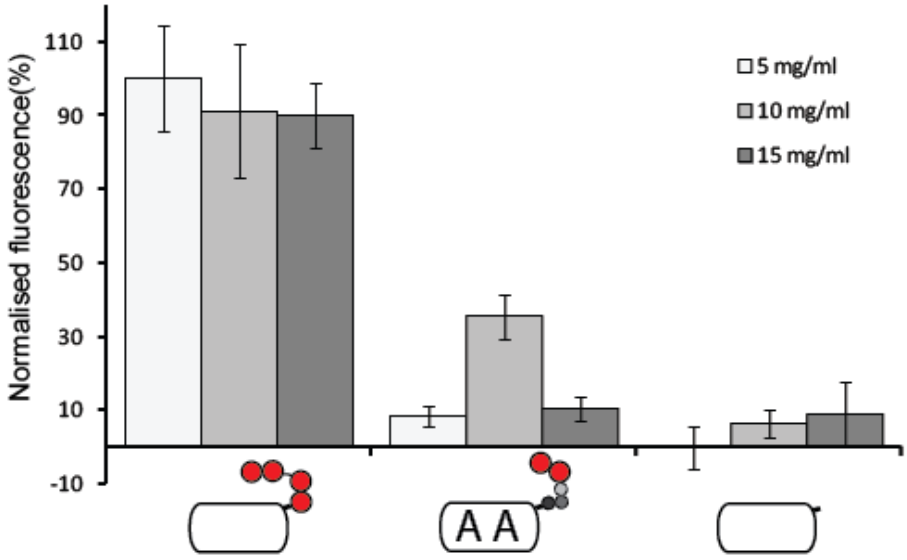


Figure 6.11 – FlAsH fluorescence of **Dps(AA)PAGCC,** which is crippled in its ability to form nanocage, and controls high totally protein concentrations in lysates. The data is normalised to the 5 mg/ml positive and negative control response. All designs examined in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) with 0.1 µM FlAsH, 3.5 mM TCEP, 1 mM EDT, 1 mM 2-ME. Lysate data from a single expression examined 12 times. Error bars are S.D. (see Figure 6.2 for explanation of schematics.)

The double mutation from arginine to alanine (R83A, R133A) has been shown to prevent nanocage formation by removing important interactions; however the small size and the neutrality of alanine might not prevent formation of the nanocage state under "forcing" conditions. Therefore it was envisioned that other mutants could be created that might prevent the nanocage from reforming through negative design. The two positions R83 and R133 were both mutated to aspartic acids (D) and tryptophan's (W) (**Dps(DD)PAGCC** and **Dps(WW)PAGCC**) with the rationale being that a charge inversion and increase in size should further cripple nanocage formation. These new control proteins were expressed, purified and characterised (see Appendix 6.3 and 6.4 for SDS-PAGE and mass spectrometry respectively).

For both proteins, no nanocage structures were observed in either SEC (Figure 6.3) or under the forcing conditions of TEM (see Appendix 6.4) and their *in vitro* fluorescence was lower than that exhibited by **DpsPAGCC** (Figure 6.4). The *in vitro* fluorescence signals for all three crippled nanocage controls displayed signals that were statistically higher than **Dps** although they had all been shown to form no nanocages. Consistent with this, when these controls were analysed via flow cytometry, populations were observed that were intermediate between negative control (**Dps**) and **DpsPAGCC**, results that were comparable to that observed for **Dps(AA)PAGCC** (Figure 6.10d). This similarity between all three dimer forming mutants in both purified and flow cytometry conditions adds confidence that the signal initially observed for **Dps(AA)PAGCC** that was higher than the **Dps** control in flow cytometry is not due to the formation of the protein nanocage state under cellular conditions, but perhaps caused by an increased background activity of FlAsH-EDT_2 in the presence of large numbers of cysteine rich proteins.

The *in vivo* conditions described above include the use of EDT to reduce background signal. While a difference between **DpsPAGCC** and **Dps(AA)PAGCC** is significant, maximizing this difference is desirable so as to more efficiently separate clones with different nanocage stabilities during eventual library screening. Therefore a range of reducing conditions were screened using flow cytometry to optimise the cytometric differentiation of the **DpsPAGCC** and the crippled nanocage clones (Figure 6.12 and Figure 6.13). We explored EDT concentrations and the use of TCEP which has been reported as an alternative reducant in cellular FlAsH conditions[8].

Increasing EDT concentration has little effect on the average fluorescence difference between bacteria expressing **DpsPAGCC** and **Dps(AA)PAGCC,** but as expected, it did slightly decrease the overall signal for both (Figure 6.13). In the presence of the reducing agent TCEP, the major population of **Dps(AA)PAGCC** clones did decrease to a larger extent than those

expressing **DpsPAGCC**, however, a second species overlapping with those expressing **DpsPAGCC** was observed (Figure 6.12). It is unclear what is the origin of these two populations is but it is clear that these conditions should be avoided. Therefore, it was decided that the best conditions involved 2 mM EDT and no TCEP. Although the close separation of these clones was a concern for a library screen, it was thought that it could be overcome by performing multiple rounds of selection.

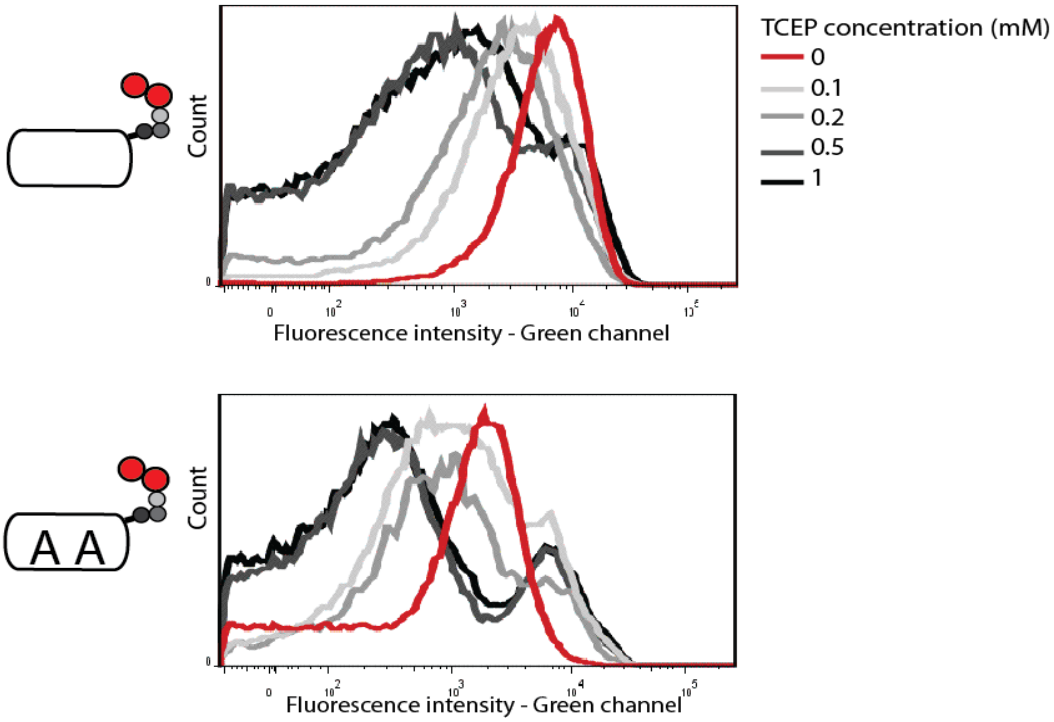


Figure 6.12 – Optimisation of reducing conditions (TCEP) for flow cytometry with cage-forming and cage-crippled clones. The green channel fluorescence intensity of **DpsPAGCC** (top) and **Dps(AA)PAGCC** (bottom) was monitored in the presence of different concentrations of TCEP. After 1.5 hours of protein expression, the cells were treated with 20 µM FlAsH-EDT_2 in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH7.8) with 2 mM EDT. After a further 1.5 hours, 0.1 µM Hoechst 33342 and 0.03 µM PI were added if indicated. Blue channel: Ex laser - 375 nm, Em filter- 450 ±20 nm. Red channel: Ex laser – 488 nm, Em filter – 670 ±20 nm. Data normalised through software (FlowJo, Tree star) using 100,000 cells.

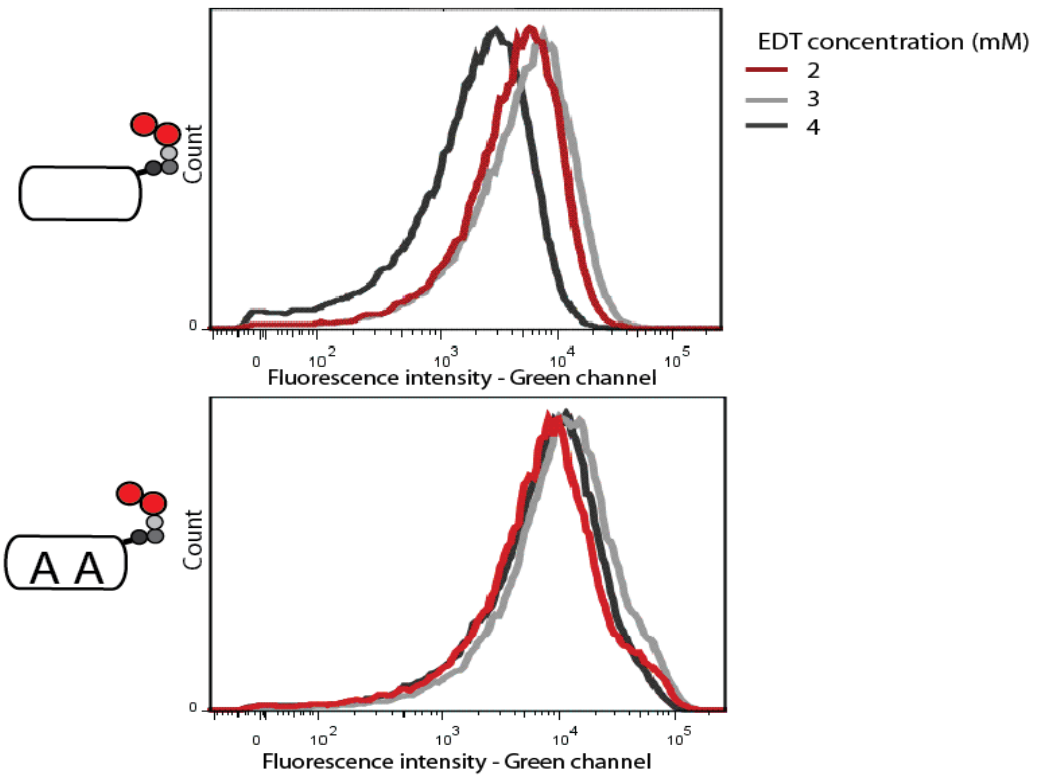


Figure 6.13 - Optimization of reducing conditions (EDT) for flow cytometry with cage-forming and cage-crippled clones. The green channel fluorescence intensity of **DpsPAGCC** (top) and **Dps(AA)PAGCC** (bottom) was monitored in the presence of different concentrations of EDT. After 1.5 hours of protein expression, the cells were treated with 20 µM FlAsH-EDT_2 in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH7.8) with 2 mM EDT. After a further 1.5 hours, 0.1 µM Hoechst 33342 and 0.03 µM PI were added if indicated. Blue channel: Ex laser - 375 nm, Em filter- 450 ±20 nm. Red channel: Ex laser – 488 nm, Em filter – 670 ±20 nm. Data normalised through software (FlowJo, Tree star) using 100,000 cells.

To summarize this section, a method to detect nanocage oligomerization states of proteins over expressed inside living *E. coli* was created. The method was characterised and verified by both flow cytometry and microscopy with large differences observed between the positive and negative controls. High fluorescence signal was observed for cells expressing the bipartite tetracysteine design, **DpsPAGCC,** and when mutations were added to disrupt nanocage formation, the fluorescence signal decreased as expected. This decrease in signal therefore permits not only the detection of self-assembly *in vivo*, expanding this technology even further, but also the ability to use cell sorting to separate bacteria expressing different levels of assembled and non-assembled proteins from each other. This will be the basis of a FlAsH/FACS, OCOB high throughput screen.

6.2.2. Development of a FlAsH/FACS, OCOB high throughput screen to recover nanocage formation from nanocage-crippled clones

Extending protein nanocage bipartite tetracysteine display technology from *in vitro* to *in vivo* monitored with flow cytometry permits the advancement to high throughput screening using cell sorting. The plate based, medium throughput *in vitro* screening described in Chapter 5 can help advance biophysical analysis, and optimise conditions for nanoparticle formation and protein stability. However, this technique has been limited to only one protein design at a time. A complementary approach therefore is required to discover new proteins with enhanced properties. Traditionally this has been done through rational design and mutagenesis of key residues followed by purification and characterisation. As stated previously, this process can be time-consuming which limits the number of designs that can be assessed. A consequence of this, is that only a narrow slice of "design space" can be explored, thus encouraging safe and conservative designs, at the expense of the risky and bold, and making the discovery of unexpected solutions to design requirements improbable. An alternative approach is to use high throughput screening so as to explore a much greater number of designs. Because purification and characterisation is often rate limiting, performing the screen inside living cells would be ideal. To further facilitate this endeavor, one clone/one bacterium (OCOB) libraries would link phenotype to genotype, providing a straight forward way to tag and identify hit clones. These libraries would be coupled to a screen dependent on FlAsH fluorescence monitored and separated by FACS. While this has been shown to be a viable strategy (see above), the last remaining hurdle is the design and production of the libraries themselves.

6.2.2.1. Library design: Dps(AA)PAGCC with a randomized protein-protein interface for the recovery of assembly

As a first generation proof that the concept of FlAsH/FACS assay for screening OBOC libraries is viable, we set out to recover assembly from nanocage-crippled clones. We used

Dps(AA)PAGCC, which is crippled in its ability to assemble into nanocages, as a background which we subjected to random mutagenesis to create a library. The aim of the research described in this section is to discover library members with recovered assembly properties. Along with establishing this technique for other, more ambitious applications, any mutants discovered in this way could shed light on the fundamentals of protein-protein interactions and quaternary structure and complement the computational interface repacking strategies that have been recently applied to protein nanocages[18].

Our strategy focused on the repacking of the protein-protein interface near the Dps-like three-fold axis of symmetry where the two residues R83 and R133 are located. Mutation of these two residues to alanine in **Dps(AA)PAGCC** causes a disruption in nanocage assemble. Positions for randomization at this interface were selected rationally by examining what amino acids interacted directly with the interfacial network made up by the R83 and R133 residues (Figure 6.14). Because of the symmetry of the homomeric protein, the impact of even a single mutation would be maximized at the three-fold interface. Residues N19, D20, V21, D141, D142, D143, S152, R153 and D156 were identified as being part of the key interfacial network (see Appendix 6.5). These residues reside at three distinct positions on the tertiary structure of the Dps monomer. The first set of residues, N19, D20 and V21, are located on the semi-unstructured N-terminal tail leading to the A-helix. The second set of residues, D141, D142 and D143 are at the N-terminus of the D-helix. Third, S152, R153 and D156 are also on the D-helix but across the interface from D141, D142, and D143 of another monomer. Further inspection of the Dps crystal structure showed that Q86, L87 and N130, while not involved directly in the interaction network, were positioned proximally in space to the network residues making them also candidates for randomization. Q86 and L87 are at the end of the C-terminal end of the B helix and N130 is positioned along the C-helix. Taken together, these represent a large proportion of the residues located at this protein-protein interface.

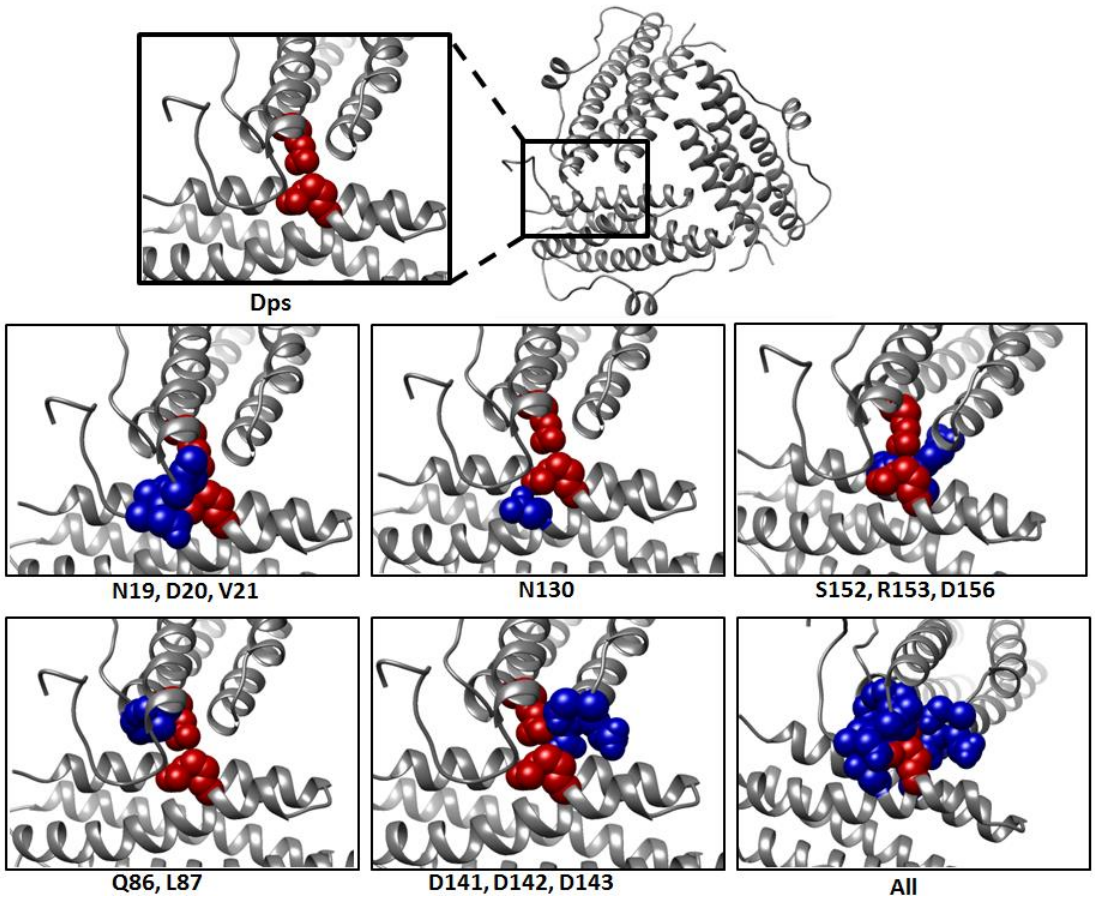


Figure 6.14 – Positions of randomization for libraries of **Dps(AA)PAGCC** in a screen to recover nanocage formation. Crystal structure is of WT Dps (PDB:1DPS) showing a protein-protein interface at the three-fold, Dps-like symmetry axis. Residues that were mutated to alanine, R83 and R133, to crippled nanocage formation are highlighted in red. The positions that were randomized in this chapter to generate **Dps(AA)PAGCC** libraries are highlighted in blue. The randomized residues are organized in sets based on primers used in the TPCR generation of the libraries.

6.2.2.2. Generation of Dps(AA)PAGCC library with TPCR

With the locations for randomization set, a method had to be employed to introduce diversity at these positions in an efficient manner. There are several methods for the creation of protein libraries including error prone PCR[19], the use of mutator strains (XL1-red) and DNA shuffling[20] (see Chapter 1). Most of these methods do not allow for the precise structural control of where points of diversity are introduced, limiting them for surgical analysis of specific protein-protein interfaces.

Transfer PCR (TPCR, see Chapter 1 and Figure 1.9) was chosen as a diversification technique as it permits the precise placement of multiple random mutations using a one pot

reaction with multiple primers[15]. Primers were designed with two different sizes: Five longer forward primers (F1, F2, F3, F4, F5) (long) that had higher melting temperatures and shorter primers (short) that had lower melting temperatures (see Appendix 6.1 for primer sequences). Codons that corresponded to the residues selected for randomization were replaced with NNS codons, where ”N” can be any natural base and “S” is restricted to C or G. This codon set limits the expression of two stop codons, and one of the codons for cysteine while including codons encoding all natural amino acids. Stop codons are avoided so as to limit truncations and cysteines were avoided to prevent artifactual FlAsH binding. It should be noted that remaining stop codons can be prevented by the use of stop suppressor bacteria strains, but that was not used for this library.

When the TPCR technique was first reported, it was very sensitive to primer concentration, and required extensive optimisation of conditions[15]. This was indeed found to be true in our hands. (see Appendix 6.6 and 6.7 for agarose gels and reaction conditions respectively). We optimized the concentration of both long and short primers in test PCRs which were analyzed by agarose gels. We found 6 nM of each of the long primers provided the highest yield of amplified plasmid, however, when this product was transformed and clones were sequenced, little incorporation of either the F1 or F5 primers was observed. Therefore new conditions used higher concentrations (10 nM) of these primers while keeping the rest at 6 nM. This led to an increase in plasmid amplification and, after sequencing, incorporation of at least four of the five of the primers (Table 6.2).

	Native mutations	Forward primer				
		F1	F2	F3	F4	F5
Dps(AA)PAGCC sequence	**83 133** **A A**	**19 20 21** **N D V**	**86 87** **Q L**	**130** **N**	**141 142 143** **D D D**	**152 153 156** **S R D**

1	A A	N D V	Q L	N	D D D	S R D
2	A A	N D V	Y V	N	D D D	S R D
3	A A	N D V	Q M	T	D D D	S R D
4	A A	N D V	Q L	N	D D D	S R D
5	A A	N D V	Q L	N	D D D	S R D
6	A A	N D V	Q L	S	D D D	S R D
7	A A	N D V	Q L	S	D D D	S R D
8	A A	N D V	R A	S	D D D	S R D
9	A A	N D V	E V	N	T L C	S R D
10	A A	N D V	F L	Y	E * D	S R D
11	A A	N D V	Y *	S	D D D	S R D
12	A A	N D V	G G	R	A G T	S R D
13	A A	N D V	Q L	L	D D D	Y V A
14	A A	N D V	Q L	N	Q L P	S R D
15	A A	N D V	Q L	N	M A A	S R D
16	A A	N D V	Q L	Y	D D D	S R D
17	A A	N D V	Q L	N	K K V	S R D
18	A A	N D V	Q L	N	D D D	S R D
19	A A	N D V	Q L	R	D D D	S R D
20	A A	N D V	Q L	D	D D D	S R D

Table 6.2 – Generation of **Dps(AA)PAGCC** library with TPCR. Sequences of twenty clones from the native TPCR library of **Dps(AA)PAGCC** with twelve sites of diversity. The degenerate residues are organized based on forward primer from which they were derived in the TPCR. Changes from **Dps(AA)PAGCC** are highlighted. *=stop. Note that a suppressor strain was not used. Grey highlighting denotes where mutations to the **Dps(AA)PAGCC** sequence are observed.

Twenty clones were sequenced to get an idea of the diversity explored in the library. (Table 6.2) It should be noted that statistically speaking, twenty clones provides no way near a significant sample of the library. However, it does provide some indication of primer incorporation, and any gross prescreening convergences of sequence could be identified. The sequencing data shows clones that have incorporated four of the five primers and some clones have multiple incorporations. However, F1 had no incorporation and F5 only appeared in one clone. Moreover, few clones had more than two or three primers incorporated.

A diverse set of amino acids were presented by the clones, however, some of the primers seem to be restricted in the residues encoded. Primer F3, for example, encodes a large number of serines at position 130. Furthermore some clones included stop codons as is

allowed by the NNS codon set (see above). These would be expected to generate truncated proteins.

After cloning, fifteen separate optimized transformations resulted in a predicted library size of 63,000 clones (4,200 transfectants per transformation). However, with twelve position of mutagenesis, this library has a potential of diversity of $4x10^{15}$ clones and clearly the actual library diversity is very far from this. Further optimization of the cloning is required to remove the background wild-type sequence and to increase the incorporation of the F1 and F5 primers. Although, complete library diversity was not achieved, 63,000 is still a large number and the library diversity was deemed large enough to attempt screening. It was felt that even if clones with no new properties were discovered, at the very least, potential pitfalls with the screen could be identified by pushing forward.

6.2.2.3. FlAsH/FACS screen of a Dps(AA)PAGCC OCOB library with a randomized protein-protein interface for the recovery of assembly

The library was screened by FACS using the three fluorophore system described above (Figure 6.9). It was sorted four times where the top 1% of most fluorescent clones were collected and amplified before the next round of sorting (Figure 6.15). The average FlAsH fluorescence of the native library was less than that for bacteria expressing **Dps(AA)PAGCC**, however this intensity increased after the first two rounds of screening. Surprisingly the fluorescence of the library after the third round was lower than that for the second round however this may be a consequence of the fact that the clones after the second round were not sorted the next day like in the other rounds but instead was kept alive for two days by diluting parts of this sample into fresh LB before being sorted. This decrease in fluorescence intensity could be attributed to a change in the state of the cells. However, the sample after the third round was again sorted for the fourth and final time. Although the general increase in fluorescence was intriguing, the decrease in fluorescence after the third round caused

concern. The clones after the fourth sort were plated and twenty of the resultant colonies were sequenced (Table 6.4).

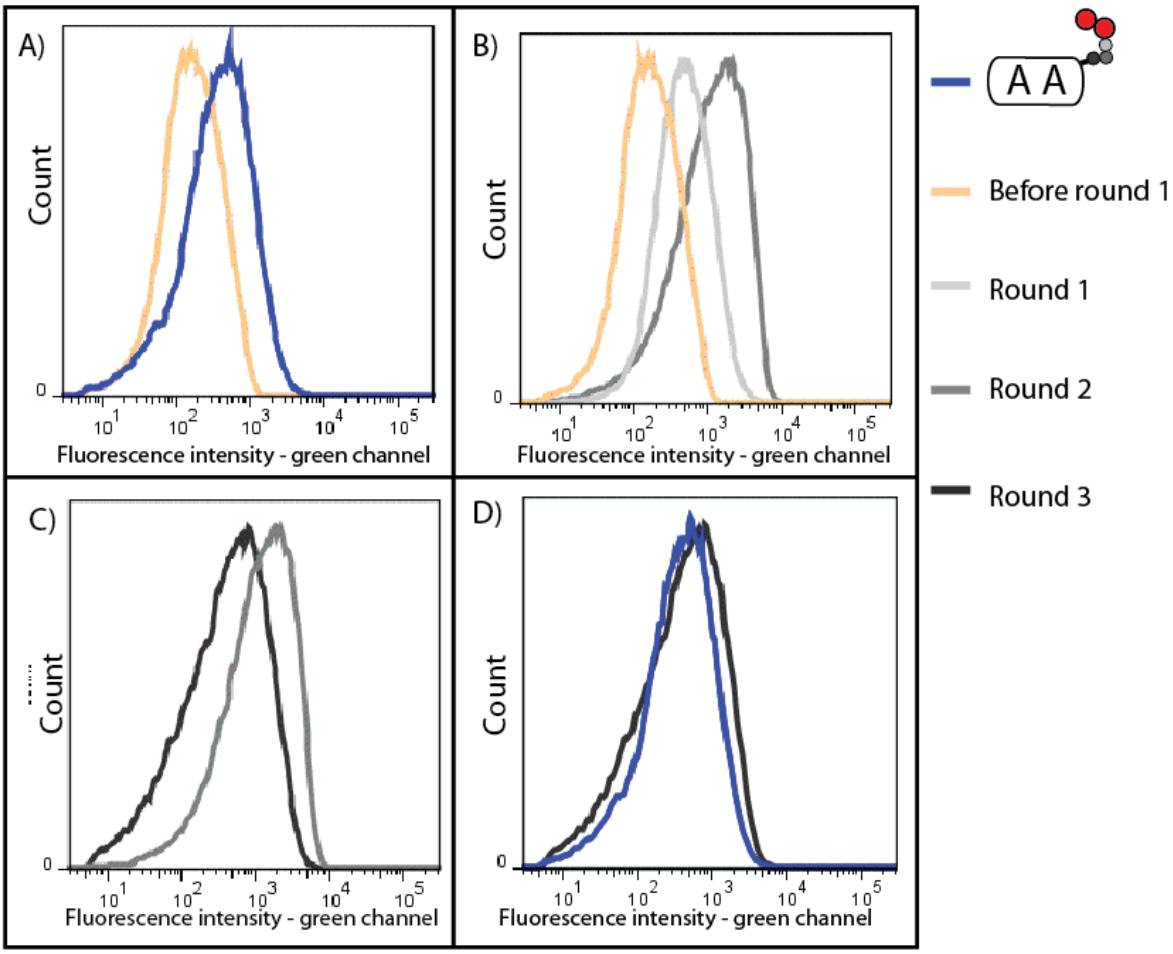


Figure 6.15 – FlAsH/FACS screen of a Dps(AA)PAGCC OCOB library with a randomized protein-protein interface for the recovery of assembly (A) Comparison of the clones expressing the naïve protein library before any selection rounds (yellow) and those expressing the cage-crippled mutant, **Dps(AA)PAGCC** (blue), (B) Comparison of clones expressing the native protein library before any selection rounds (yellow), after one round of selection (light grey) and after two rounds (grey). (C) Comparison of clones expressing the library after two rounds of selection (grey) and after three rounds (dark grey). Note that the later sample was amplified for two days instead of the usual one day before the round of selection., (D) Comparison of clones expressing the library after three rounds of selection (dark grey) and those expressing the nanocage-crippled mutant, **Dps(AA)PAGCC** (blue). After 1.5 hours of protein expression, the cells were treated with 20 µM FlAsH-EDT_2 in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH7.8) with 2 mM EDT. After a further 1.5 hours, 0.1 µM Hoechst 33342 and 0.03 µM PI were added if indicated. Blue channel: Ex laser - 375 nm, Em filter- 450 ±20 nm. Red channel: Ex laser – 488 nm, Em filter – 670 ±20 nm. Data normalised through software (FlowJo, Tree star). During a round of selection, the top 1% were collected and amplified for either 12 hours (rounds 1,2,3) or 48 hours (round 4) before being subjected to the next round.

Sequencing of the twenty clones after four rounds of FlAsH/FACS seemed to indication some trends, however, it should be noted that the small number of clones

sequenced of the native library and the mature library precludes anything but superficial analysis of changes in the library composition or the absence of any specific clones. However, we hope that after more thorough optimisation to increase diversity, deep sequencing techniques will be used to follow the maturation of the library with increasing selection pressure.

	Native mutations	Forward primer F1	F2	F3	F4	F5
Dps(AA)PAGCC sequence	**83 133** **A A**	**19 20 21** **N D V**	**86 87** **Q L**	**130** **N**	**141 142 143** **D D D**	**152 153 156** **S R D**
1 #	A A	N D V	V F	F	D D D	S R D
2 #	A A	N D V	R G	*	D D D	S R D
3	A A	N D V	Q L	V	D D D	S R D
4 #	A A	N D V	Q L	Y	D D D	S R D
5	A A	N D V	Q L	V	D D D	S R D
6	Deletion					
7	A A	N D V	* Q	N	D D D	R E I
8	A A	N D V	Q L	Y	D D D	S R D
9	A A	N D V	Q L	Y	D D D	S R D
10	A A	N D V	Q L	Y	D D D	S R D
11	A A	N D V	Q L	Y	D D D	S R D
12	A A	N D V	Q L	P	D D D	S R D
13	A A	N D V	Q L	N	D D D	S R D
14	A A	* T M	Q L	S	D D D	S R D
15 #	A A	N D V	Q L	L	D W G	S R D
16	A A	N D V	Q L	Y	D D D	S R D
17	A A	N D V	Q L	P	D D D	S R D
18	A A	N D V	Q L	N	D D D	S R D
19	A A	N D V	Q L	L	D D D	S R D
20	A A	N D V	Q L	P	D D D	S R D

Table 6.4 – Sequencing results of clones after FlAsH/FACS screen. Sequences of twenty clones from the screened library after four rounds of selection. The degenerate residues are organized based on forward primers from which they were derived in the TPCR. Changes from **Dps(AA)PAGCC** are highlighted. *=stop. Note that a suppressor strain was not used. #=sequences that were taken for purification and characterisation. Grey highlighting denotes where mutations to the **Dps(AA)PAGCC** sequence are observed.

6.2.2.4. Characterisation of Dps(AA)PAGCC mutants obtained from FlAsH/FACS high throughput screen

Of the twenty sequences cloned from the library after four rounds of selection (Table 6.4), the most commonly observed mutant was **Dps(AA)PAGCC^{N130Y}** which was present in six of the sequenced clones. Although initial analysis of the native library (see Table 6.2) suggested that the TPCR method favoured incorporation of the F3 primer, the matured library exhibited an enhanced percentage of sequences with mutations at position 130, suggesting that this is a good site from which to stabilise nanocage formation. Therefore, it was decided to carry this protein on to the characterisation step. Only two sequences were observed with mutations arising from more than one included TPCR primer and no stop codon: **Dps(AA)PAGCC^{Q86V,L87F,N130F}** and **Dps(AA)PAGCC^{N130L,D141D,D142W,D143G}**. These two proteins were also carried to the next step. One of the most striking aspects of the sequences of these twenty clones is the high percentage containing stop codons, which would not be a concern if using suppressor strains. Therefore **Dps(AA)PAGCC^{Q86V,L87G,N130Stop}** was also carried through to confirm that it was indeed producing a truncated protein. These four proteins were expressed, purified, and characterised (see Appendices 6.2 and 6.3 for SDS-PAGE and mass spectrometry respectively). It was confirmed by mass spectrometry that **Dps(AA)PAGCC^{Q86V,L87G,N130Stop}** was indeed truncated. Why we seem to be selecting for truncated proteins is currently unknown. It should be emphasized that the truncated proteins are not expressing the PAGCC and therefore should not have a FlAsH binding site even if assembled.

The three full-length proteins were assayed for their FlAsH-dependent, *in vitro* fluorescence as described in Chapters 4 and 5 (Figure 6.16). While their signals were significantly larger than the crippled nanocage mutant **Dps(AA)PAGCC**, they were not a high as the nanocage forming protein **DpsPAGCC**.

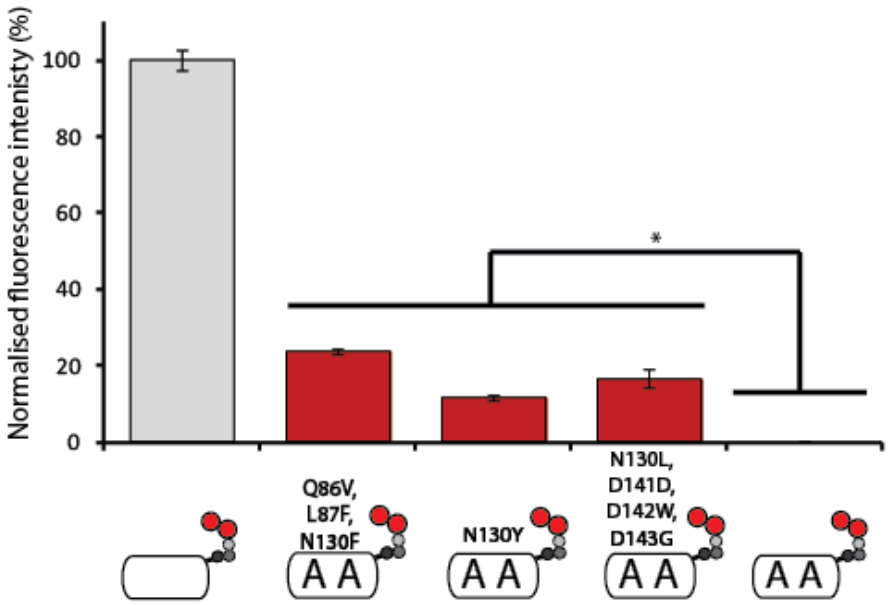


Figure 6.16 - FlAsH fluorescence of purified proteins from FlAsH/FACS screening (see Figure 6.2 for explanation of schematics). All proteins are 0.1 mg/ml as purified proteins, in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) with 0.1 µM FlAsH, 3.5 mM TCEP, 1 mM EDT, 1 mM 2-ME. * Two-tailed P-values = 0.0001. The data is normalised to **DpsPAGCC** and **Dps(AA)PAGCC**. The data is averaged from 8 replicates. Error bars are S.D.

Each protein was further characterized with SEC and TEM. (Figure 6.17 for SEC and see appendix 6.4 for TEM) For the control **Dps(AA)PAGCC**, neither TEM or SEC showed any nanocage-like structures. All three proteins obtained from the screen to recover nanocage-assembly showed nanocage structures in TEM suggesting that the screen was successful. This was further emphasized by the presence of nanocage peaks in the SEC for all three proteins. In the SEC, **Dps(AA)PAGCC**^{Q86V,L87F,N130F} and **Dps(AA)PAGCC**^{N130L,D141D,D142W,D143G}, showed a larger ratio of nanocage than **Dps(AA)PAGCC**^{N130Y}**,** however the dimer species was still the dominant population for all three. Interestingly, an intermediate species seems to exist, generating a peak between the nanocage and dimer. Although this is speculative, this result raises the possiblity that we are stablizing an intermediate that isn't nanocage but that also has a similar three-fold interface. This should not be a complete surprise because, in retrospect, the screen as designed is screening for the formation of a three-fold symmetry axis and not explicitly the nanocage. Although the implications of this have not been fully realized, this might be an important lesson to keep in mind.

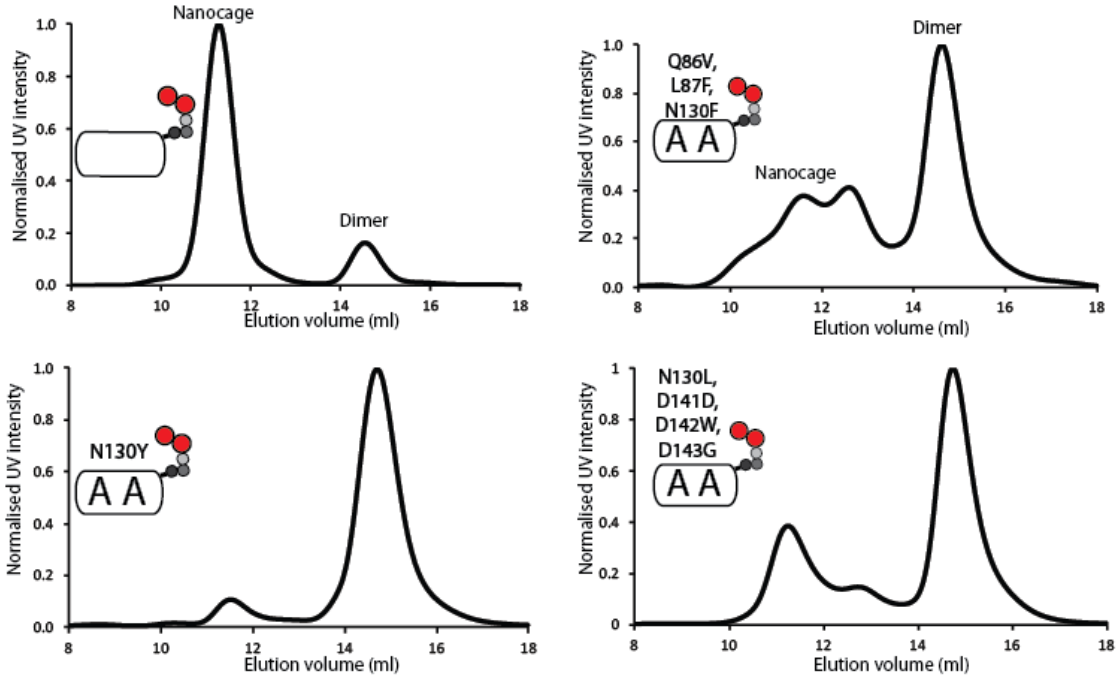


Figure 6.17 – Size exclusion chromatograms of purified proteins from FlAsH/FACS cage recovery screening. Traces are averages of three runs, normalised to the highest intensities for each (0.5 ml injection of 0.5 mg/ml protein). For explanation of cage-crippled mutants see Section 6.2.1.3. See Figure 6.2 for explanation of schematics.

To further confirm the recovered nanocage forming ability of these proteins, they were further characterised by dynamic light scattering (DLS) (Figure 6.18). All three crippled nanocage controls, (**Dps(AA)PAGCC**, **Dps(DD)PAGCC** and **Dps(WW)PAGCC**) exhibited a small hydrodynamic diameter while all nanocage forming controls, **DpsCCPGCC**, **DpsPAGCC** and **Dps** assembled into larger structures. Further confirming that the screen was successful, all three mutants from the nanocage recovery screen, **Dps(AA)PAGCC$^{Q86V,\ L87F,\ N130F}$**, **Dps(AA)PAGCC$^{N130L,\ D141D,\ D142W,\ D143G}$**, and **Dps(AA)PAGCC^{N130Y},** exhibited hydrodynamic diameters on par with the nanocage-forming controls.

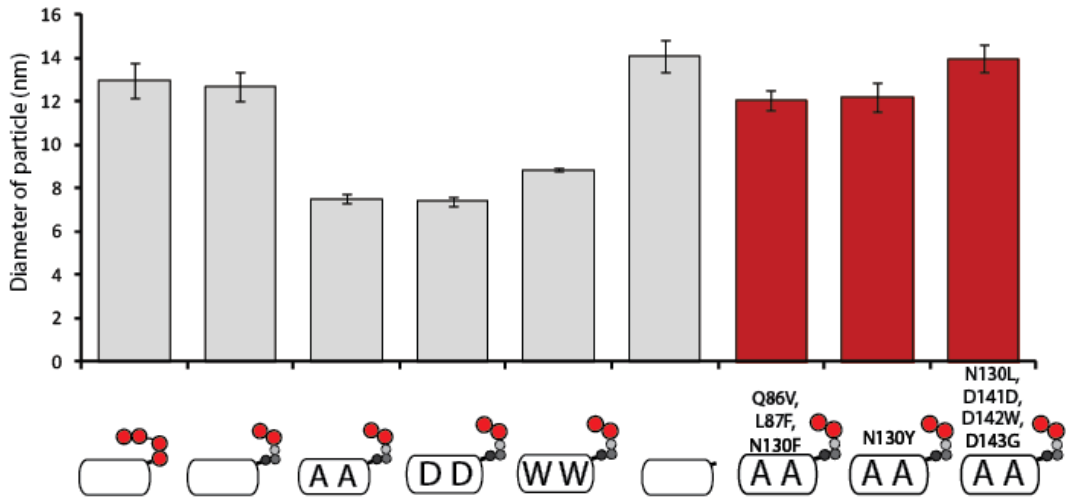


Figure 6.18– Dynamic light scattering analysis of purified proteins from FlAsH/FACS cage recovery screen. Each protein was 0.2 mg/ml in FLAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8). Data is the average of three replicates. Error bars are S. D. See Figure 6.2 for explanation of schematics.

Taken together, the DLS, SEC and TEM data, point to the conclusion that **Dps(AA)PAGCC**$^{Q86V,\ L87F,\ N130F}$, **Dps(AA)PAGCC**$^{N130L,\ D141D,\ D142W,\ D143G}$, and **Dps(AA)PAGCC**^{N130Y} can all assemble into the nanocage oligomerization state and that the mutations, discovered by our high throughput screen can recover assembly even in the presence of the nanocage crippling mutations R83A and R133A. These three mutants represent 40% of the sequences obtained from the twenty sequenced clones from the mature library suggesting that this technology, even at an unoptimized, proof-of-concept stage, can discover new nanocages with novel assembly properties. Although nanocages were observable the ratio of nanocage to dimer was not recovered to the distribution observed for **DpsPAGCC**.

Along with nanocage stability, the secondary structure and thermal stability of the proteins isolated from the screen was also investigated. This data (see Appendices 6.8, 6.9 and 6.10) demonstrated that the proteins cooperatively denature indicating that they are well folded, however, the melting points and isotherm shapes are distinct. **DpsPAGCC** exhibits a single melting transition (67 ^{0}C) while **Dps(AA)PAGCC** displays two transitions (43 ^{0}C and 73 ^{0}C) indicating, multiple states, or a longer lived folding intermediate. Two mutants derived from

the screens had similar bimodal isotherm shape to that of their parent with a low temperature transition near 43 ^{0}C. The lack of ability of the mature library mutants to overcome the change in thermal stability caused by the assembly crippling mutations could be indicative of the extra intermediates observed in SEC (Figure 6.17).

It should be noted that this screen was decidedly focused on discovering proteins with the enhanced ability to assemble into nanocages. It was not designed to discover proteins with enhanced thermal stability of secondary structure. When working with these and similar protein nanocages[17, 21, 1], we have found that these two properties, assembly and thermal stability, while surely coupled in some ways, are not always congruent. So it should not come as a complete surprise that thermal stability does not always scale with assembly.

The appearance of a possible assembly intermediate is interesting. Although much more characterisation is required, its presence may suggest a non-native mechanism to assembly. This also is not surprising because, the design rules of the library, retained the R83A, R133A mutation, and randomised the rest of the interface. It stands to reason that any nanocage forming clones would be required to converge on a solution to assembly that is an alternative one to the wild-type protein.

While we have been successful in designing a high throughput screen to recover nanocage assembly in a nanocage-crippled protein, clearly there is much room for enhancement. One of the clear weaknesses is the poor coverage of library diversity. The TPCR strategy needs further optimization to ensure robust incorporation of all five primers. In addition a simple way to expand diversity would be to increase the number of transformations from fifteen to one-hundred. While this is somewhat crude, it would increase the number of diverse clones from 63,000 to nearly 630,00 and could easily be accomplished in an afternoon without any additional optimisation. Furthermore, a mutagenesis step between screenings around could easily be incorporated. Ways to increase stringency or to more strongly couple

nanocage formation with thermal stability would be to perform expression or FlAsH treatment (30 °C and 25 °C respectively) at higher temperatures, although it should be noted that protein expression, cell metabolism, etc would all be affected. Finally, as we think that we are may be seeing new intermediates that have nanocage-like symmetry (see the discussion about the SEC data above), perhaps employing alternative FlAsH binding sites or even multiple types in concert should be explored.

6.3. Conclusions

In summary, the research described in this chapter expanded the *in vitro* FlAsH-based assay for the direct detection of Dps assembly described in Chapter 5 into living bacteria and a high throughput screen using one clone/one bacterium libraries sorted with FACS. As far as we know, this is the first example of a high throughput screen for protein nanocage engineering and the first application of a FlAsH-based OC/OB screen using FACS.

This *in vivo* method was confirmed with both fluorescence microscopy and flow cytometry. Clones expressing negative and positive control proteins demonstrated very low and very high fluorescence respectively and cells expressing the protein with the bipartite FlAsH binding site developed in Chapter 5 (**DpsPAGCC**) exhibited intermediate fluorescence. The technique was more deeply explored using controls that were known to have crippled assembly properties and cells expressing proteins with these mutations applied to DpsPAGCC were observed to have less fluorescence.

The *in vivo* assay to determine when a protein does or does not self-assemble into a nanocage was then developed into part of a high throughput screen to recover assembly from a nanocage-crippled background by protein-protein interface repacking. A cloning method was developed to introduce random mutations at twelve specific locations in the Dps structure which were located at the same interface near the nanocage-crippling mutations. The resulting library was transformed into cells in a one clone/one bacterium manner allowing each cell to express a different clone and providing the sequencing tag to identify the clone. The bacteria with the strongest FlAsH fluorescence were collected with FACS. After four rounds of selection, twenty clones were sequenced. Of these, three mutant proteins were purified and characterised with SEC, TEM, DLS, and CD. All three mutants could assemble into nanocage structures despite containing the mutations that were assembly crippling in the parent protein. Thus, this study successfully establishes a first generation high throughput screen to

discover protein nanocages with enhanced assembly properties applied to protein-protein interface repacking. We think that this high throughput technique could be used to discover better FlAsH binding sites (i.e. to enhance the technique itself), to understand more deeply the role of protein-protein interactions in nanocage assembly, complimenting computational strategies (as discussed in Chapter 1), and to identify protein nanocages with novel properties and, perhaps, structures.

6.4. Methods and Materials

6.4.1. Cloning

6.4.1.1. Transfer of genes developed in previous chapters to pET-46 (DpsCCPGCC, DpsPAGCC and Dps)

The plasmids containing the genes coding for **DpsCCPGCC**, **Dps** and **DpsPAGCC** were transferred to pET-46 via amplification with primers that provided ligation independent cloning (LIC) sites[4]. The PCR solution (Pfu reaction buffer (Promega, 5 µl of 10x), dNTP mix (Promega, 2 µl of a solution containing dATP, dTTP, dGTP, and dCTP at 2 mM each), forward and reverse primers (400 ng each, IDT, Germany), template (100 ng) and Pfu polymerase (Promega, 2 µl of 2.5 U/µl) in 50 µl total volume) was subjected to an initial melting step (95 °C for 30 s), followed by 30 cycles of amplification (95 °C for 30 s, 58 °C for 45 s and 72 °C for 1 min) followed by 72 °C for 7 min. The resulting PCR product was isolated by gel purification (Promega). The PCR product was then treated with T4 polymerase in order to create the required complementary overhangs (NEB buffer 2, 2 µl of 10x buffer), dATP (NEB, 2 µl of 25 mM), BSA (NEB, 0.2 µl of 100x), PCR product (0.3 pmol) and T4 DNA polymerase (NEB, 0.6 µl of 10 U/µl) in a total volume of 20 µl 30 min, 22 °C). The insert was annealed to open plasmid pET-46 (Novagen) (1 µl of the T4 Pol product, 0.5µl of plasmid solution (50 ng/µl)) by incubating (room temperature, 30 min) followed by a second incubation (room temperature, 30 min) with EDTA (1 µl of 100 mM). The constructs (2.5 µl) were transformed (Xl-1 blue, Novagen) and the resulting colonies were assessed by colony PCR, a solution of Gotaq buffer (Promega, 5 µl of 5x), dNTP mix (Promega, 2 µl of a solution containing dATP, dTTP, dGTP, and dCTP at 2 mM each), T7 promoter and terminator primers (150 ng each, Eurofins MWG), taq DNA polymerase (Promega, 0.25 µl of 5 U/ µl), colony suspension (10 µl of a 50 µl total colony suspension in deionised water) and deionised water (total volume of 25 µl), was subjected to an initial melting step (95 °C for 10 min) followed by 30 cycles of amplification, ((95 °C for 30 s, 59 °C for 30 s and 72 °C for 30 s) followed by 72 °C for 7 min). Clones demonstrating a PCR

product of the predicted size were subjected to miniprep (Sigma) and the resulting purified plasmids were sequenced (see Sequencing Appendix 6.1 for sequencing results and Appendix 6.1 primers).

6.1.1.2. Introduction of mutations for nanocage-crippled mutants Dps(AA)PAGCC, Dps(DD)PAGCC and Dps(WW)PAGCC

Using **DpsPAGCC** in pET-46 as a template (see section 6.1.1.1), site directed mutagenesis was performed to create controls that only exist in the dimer oligomerization state. The reaction mixture containing Pfu buffer (Promega, 2 µl of 10x), dNTP mix (Promega, 2 µl of a solution containing dATP, dTTP, dGTP, and dCTP at 2 mM each), forward and reverse primers (125 ng each, IDT, Germany), DMSO (1% V/V), the dsDNA template (pET-46 designs, 100 ng) and Pfu polymerase (Promega, 1 µl of 3 U/µl) in a total volume of 25 µl was subjected to melting (95 °C for 5 min), followed by 20 cycles of amplification (95 °C for 1 min, 65 °C for 1 min and 72 °C for 6 min) followed by 72 °C for 10 min. The reaction was then digested with Dpn1 (NEB buffer 4 (1 µl of 10x), amplification product (5 µl), deionized water (3.5 µl), Dpn1 (NEB, 0.7 µl of 10 U/µl), 37 °C, 2h). The resulting solution was transformed (XL-1 blue, Novagen) and the subsequent colonies subjected to miniprep (Sigma) and sequenced (see Sequencing Appendix 6 for sequencing results and Appendix 6.1 primers).

6.1.1.3. TPCR library generation

The TPCR library used **Dps(AA)PAGCC** in pET-46 as a template and included 5 forward primers, with randomized residues, and 1 reverse. The PCR reaction contained a total volume of 50 µl including, Pfu reaction buffer (Promega, 5 µl of 10), dNTP mix (Promega, 2 µl of a solution containing dATP, dTTP, dGTP, and dCTP at 2 mM each), forward and reverse primers (8 nM of F1 and F5, 6 nM of F2, F3, F4 and R1, IDT, Germany) dsDNA template (20 ng), DMSO (2% V/V), Pfu polymerase (promega, 1 µl of 3 U/µl) and deionised H_2O (to a total of 50 µl) which was subjected to an initial melting step (95 °C for 5 min) followed by 40 cycles of amplification

(95 °C for 30 s, 58 °C for 1 min, raising 0.3 °C every cycle and 72 °C for 6 min) followed by 72 °C for 10 min. The product of this reaction was then digested with Dpn1 in a reaction mixture containing buffer (NEB buffer 4 (2 µl of 10x), amplification product (17 µl), Dpn1 (NEB, 1 µl of 10 U/µl), 37 °C, 1.5 h). The solution was desalted (Promega, PCR clean up wizard) before any transformations (see Sequencing Appendix 6 for sequencing results and Appendix 6.1 primers).

6.4.2. *In vivo* analysis

6.4.2.1. Live cell fluorescence microscopy

E.coli containing Dps variants in the vector pET-46, were grown from frozen stocks in LB (5 ml, 37 ^{0}C, overnight) with 1% being added to fresh LB (5 ml) and grown (37 °C) until an O.D$_{600}$ of 0.6. Protein expression was then induced by the addition of IPTG (2.5 µl of a 1 M stock) and the cultures were further incubated (1.5 h, 30 °C). From this culture, 100 µl was taken and centrifuged (4000 rpm, 30 s) and resuspended ((20 µl of FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) with EDT (final concentration of 2 mM)) to which FLAsH-EDT$_2$ was added (Invitrogen, 0.2 µl of 2 mM stock) followed by further incubation (25 °C, 1.5 h). This solution was first pelleting with centrifugation (4000 rpm, 30 s) then washed twice (using 20 µl of FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) with EDT (final concentration of 2 mM)) followed by a final re-suspension (100 µl, FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) and kept at 4 °C). To this solution, the fluorophores needed to screen for bacterial survival were added (Invitrogen, Hoechst 33342, 1 µl of a 5 mg/ml stock and Propidium iodide, 2 µl of a 1 mg/ml stock). To a polylysine slide, 5 µl of the cells was added with a cover slip placed on top and left at 4 °C. Images were taken using a wide field Nikon Eclipse Ti-E Inverted microscope at 100x magnification (Filter sets, Hoechst 33342 Ex - 402 ±15 nm, Em - 455 ±50 nm. FlAsH Ex – 490 ±20 nm, Em – 525 ±36 nm. Propidium iodide Ex – 555 ±25 nm, Em – 605 ±52 nm) with images analysed using ImageJ and NIC element software.

6.4.2.2. Flow cytometry analysis

E.coli containing Dps variants in the vector pET-46, were grown from frozen stocks in LB (5 ml, 37 °C, overnight) with 1% being added to fresh LB (5 ml) and grown (37 °C) until an $O.D_{600}$ of 0.6. Protein expression was then induced by the addition of IPTG (2.5 µl of a 1 M stock) and the cultures were further incubated (1.5 h, 30 °C). From this culture, 100 µl was taken and centrifuged (4000 rpm, 30 s) and resuspended ((20 µl of FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) with EDT (final concentration of 2 mM)) to which FLAsH-EDT_2 was added (Invitrogen, 0.2 µl of 2 mM stock) followed by further incubation (25 °C, 1.5 h). This solution first pelleting with centrifugation (4000 rpm, 30 s) and then washed twice ((using 20 µl of FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) with EDT (final concentration of 2 mM)) and then resuspended (1 ml, FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) and kept at 4 °C). To this solution, the fluorophores needed to screen for bacterial survival were added (Invitrogen, Hoechst 33342, 1 µl of a 5 mg/ml stock and PI, 2 µl of a 1 mg/ml stock). This solution was filtered (DB, flow cytometry tube) and run on a DB Fortessa (Hoechst 33342 Ex laser - 375 nm, Em filter - 450 ±20 nm. FlAsH Ex laser – 488 nm, Em filter – 530 ±30 nm. Propidium iodide Ex laser – 488 nm, Em filter – 670 ±20 nm) with data being analysed on FACSDiva and FlowJo (Tree star) software. Gating was first applied to the $Hoechst^+$ response to collect all bacterial cells. This population was further gated to collect all PI^- (alive) cells, which was lastly analysed for FlAsH fluorescence.

6.4.2.3. FlAsH/FACS of one clone/one bacterium libraries

The desalted constructs (8 µl) were electroporated (XL-1 Blue, Novagen) and recovered using SOC medium and left to incubate (overnight, 37 °C) and inoculated (carbenicillin, 1 µl of a 50 mg/ml stock per 1 ml of SOC). The plasmids were collected via miniprep (Sigma) and re-electroporated (8 µl (Rosetta, Novagen)) using SOC medium to recover and was then left to incubate (overnight, 37 ^{0}C) and inoculated (carbenicillin, 1 µl of a

50 mg/ml stock per 1 ml of SOC). This process was repeated 15 times to create the library described here.

The bacteria was pelleted via centrifugation (3500 rpm, 10 min) the SOC media was removed and fresh LB was added and inoculated (carbenicillin, 1 µl of a 50 mg/ml stock per 1 ml of LB) and left to incubate (1 h, 37 °C) followed by the induction of protein expression (IPTG, 0.5 µl of 1 M stock for each 1 ml of LB, 30 °C, 1.5 h). For the first round, 4 x 1 ml samples were taken and centrifuged (4000 rpm, 1 min) and resuspended ((100 µl of FlAsH buffer (100 mM Tris.HCL, 100 mM NaCl, 1 mM EDTA, pH 7.8) with EDT (final concentration of 2 mM)) and FLAsH-EDT_2 was added (Invitrogen, 1 µl of 2 mM stock) followed by further incubation (25 °C, 1.5 h). This solution first pelleted with centrifugation (4000 rpm, 30 s) and washed twice ((using 100 µl of FlAsH buffer (100 mM Tris.HCL, 100 mM NaCl, 1 mM EDTA, pH 7.8) with EDT (final concentration of 2 mM)) and then re-suspension (1 ml, FlAsH buffer (100 mM Tris.HCL, 100 mM NaCl, 1 mM EDTA, pH 7.8) and kept at 4 °C). To this solution, the fluorophores need to screen for bacterial survival were added (Invitrogen, Hoechst 33342, 1 µl of a 5 mg/ml stock and PI, 2 µl of a 1 mg/ml stock). This solution was filtered (DB, flow cytometry tube) and run on a DB Aria III (Hoechst 33342 Ex laser - 375 nm, Em filter - 450 ±20 nm. FlAsH Ex laser – 488 nm, Em filter – 530 ±30 nm. Propidium iodide Ex laser – 488 nm, Em filter – 670 ±20 nm) with data being analysed on FACSDiva and FlowJo (Tree star) software. Gating was first applied to the $Hoechst^+$ response to collect all bacterial cells. This population was further gated to collect all PI^- (alive) cells, which was lastly analysed for FlAsH fluorescence with the highest intensity cells being collected. The collected cells were placed into inoculated LB (total of 10 ml, with 10 µl of carbenicillin 1 M stock) and left to incubate (overnight, 37 °C). The process described above was repeated for each round of sorting. After the final sort, the solution was plated on LB agar plates (50 µl/ml of carbenicillin and 34 µl/ml of chloramphenicol) with resulting colonies submitted for sequencing.

6.4.3. Protein characterisation

6.4.3.1. *In vitro* protein FlAsH analysis

The pET-46 vectors containing the Dps variants were transformed into Rosetta *E. coli* cells (Novagen) and plated on LB agar plates (50 µl/ml of carbenicillin and 34 µl/ml of chloramphenicol). Selected colonies were then grown in LB (9 ml, 37 °C, overnight) as a pre-culture which was added to LB (900 ml) and grown (37 °C) until an $O.D_{600}$ of 0.6. Protein expression was then induced by the addition of IPTG (400 µl of a 1 M stock) and the cultures were further incubated (3 h, 30 °C). The cells were isolated by centrifugation (4,000 rpm, 20 min, 4 °C). The cell pellet was resuspended in lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 40 mM Imidazole, 1 mM EDTA, pH 8) and sonicated (Misonix, ultrasonic cell distruptor, pulsed 5 s on, 5 s off for 5 min). The protein solution was clarified by centrifugation (15,000 rpm, 45 min at 4 °C) and then filtered (Sartorius, 0.2 µm). The protein was purified via affinity purification (GE, Histrap FF, 5 ml, (wash buffer-40 mM Imidazole, 50 mM NaH_2PO_4, 300 mM NaCl, pH 7.4), (elution buffer-500 mM Imidazole, 50 mM NaH_2PO_4, 300 mM NaCl, pH 7.4)). Digestion to remove the short His_6 tag was not performed on any pET-46 variants. The protein solution was further purified by size exclusion chromatography (GE Hiload 16/60 Superdex, running buffer-FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8).

In vitro analysis of all Dps variants were performed as described in **5.4.3.** in the Corning black 96 well plate format.

6.4.3.1. Analytical Size Exclusion Chromatography (SEC)

The samples (0.5 ml of 1 mg/ml) in FlAsH buffer (100 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) were injected onto the column (GE Superdex 200 increase) at 1 ml/min with each protein repeated 3 times. The column was calibrated using six proteins as standards (GE Biosystems Calibration Kit).

6.4.3.2. Transmission Electron Microscopy (TEM)

TEM was performed as described in **5.4.6.2.** (See Appendix 6.4)[16].

6.4.3.3. Mass Spectrometry

Mass spectrometry data was obtained as described in **4.4.4.5.** (See Appendix 6.3).

6.4.3.4. Circular Dichorism (CD)

Circular Dichorism data was obtained as described in **5.4.6.3.** (See Appendix 6.8, 6.9 and 6.10)[22].

6.4.3.5. Dynamic light scattering (DLS)

Purified proteins were diluted to a concentration of 0.2 mg/ml in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) and analysed in a 1 cm path length cuvette on a Malvern Zetasizer nanoseries, with each protein repeated 3 times.

6.5. References

1. Zhang, Y.; Fu, J.; Chee, S. Y.; Ang, E. X. W.; Orner, B. P., Rational disruption of the oligomerization of the mini-ferritin E. coli DPS through protein-protein interface mutation. *Protein Science* **2011,** *20* (11), 1907-1917.
2. Zhang, Y.; Orner, B. P., Self-Assembly in the Ferritin Nano-Cage Protein Superfamily. *International Journal of Molecular Sciences* **2011,** *12* (8), 5406-5421.
3. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **2004,** *25* (13), 1605-1612.
4. Cornell, T. A.; Fu, J.; Newland, S. H.; Orner, B. P., Detection of Specific Protein–Protein Interactions in Nanocages by Engineering Bipartite FlAsH Binding Sites. *Journal of the American Chemical Society* **2013,** *135* (44), 16618-16624.
5. Griffin, B. A.; Adams, S. R.; Tsien, R. Y., Specific covalent labeling of recombinant protein molecules inside live cells. *Science* **1998,** *281* (5374), 269-272.
6. Griffin, B. A.; Adams, S. R.; Jones, J.; Tsien, R. Y., Fluorescent labeling of recombinant proteins in living cells with FlAsH. *Applications of Chimeric Genes and Hybrid Proteins Pt B* **2000,** *327*, 565-578.
7. Zhang, J.; Campbell, R. E.; Ting, A. Y.; Tsien, R. Y., Creating new fluorescent probes for cell biology. *Nature Reviews Molecular Cell Biology* **2002,** *3* (12), 906-918.
8. Enninga, J.; Mounier, J.; Sansonetti, P.; Van Nhieu, G. T., Secretion of type III effectors into host cells in real time. *Nature Methods* **2005,** *2* (12), 959-965.
9. Ignatova, Z.; Gierasch, L. M., Monitoring protein stability and aggregation in vivo by real-time fluorescent labeling. *Proceedings of the National Academy of Sciences of the United States of America* **2004,** *101* (2), 523-528.
10. Romantsov, T.; Helbig, S.; Culham, D. E.; Gill, C.; Stalker, L.; Wood, J. M., Cardiolipin promotes polar localization of osmosensory transporter ProP in Escherichia coli. *Molecular Microbiology* **2007,** *64* (6), 1455-1465.
11. Romantsov, T.; Battle, A. R.; Hendel, J. L.; Martinac, B.; Wood, J. M., Protein Localization in Escherichia coli Cells: Comparison of the Cytoplasmic Membrane Proteins ProP, LacY, ProW, AqpZ, MscS, and MscL. *Journal of Bacteriology* **2010,** *192* (4), 912-924.
12. Leung, D. W., E. Chen, and D. V. Goeddel., A method for the random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique* **1989,** (1), 11-15.
13. Parikh, A.; Guengerich, F. P., Random mutagenesis by whole-plasmid PCR amplification. *Biotechniques* **1998,** *24* (3), 428-431.
14. Miyazaki, K.; Arnold, F. H., Exploring nonnatural evolutionary pathways by saturation mutagenesis: Rapid improvement of protein function. *Journal of Molecular Evolution* **1999,** *49* (6), 716-720.
15. Erijman, A.; Dantes, A.; Bernheim, R.; Shifman, J. M.; Peleg, Y., Transfer-PCR (TPCR): A highway for DNA cloning and protein engineering. *Journal of Structural Biology* **2011,** *175* (2), 171-177.
16. Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W., NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **2012,** *9* (7), 671-675.
17. Fan, R.; Boyle, A. L.; Cheong, V. V.; Ng, S. L.; Orner, B. P., A Helix Swapping Study of Two Protein Cages. *Biochemistry* **2009,** *48* (24), 5623-5630.
18. King, N. P.; Sheffler, W.; Sawaya, M. R.; Vollmar, B. S.; Sumida, J. P.; André, I.; Gonen, T.; Yeates, T. O.; Baker, D., Computational Design of Self-Assembling Protein Nanomaterials with Atomic Level Accuracy. *Science* **2012,** *336* (6085), 1171-1174.

19. Chen, M. M. Y.; Snow, C. D.; Vizcarra, C. L.; Mayo, S. L.; Arnold, F. H., Comparison of random mutagenesis and semi-rational designed libraries for improved cytochrome P450 BM3-catalyzed hydroxylation of small alkanes. *Protein Engineering Design and Selection* **2012,** *25* (4), 171-178.
20. Stemmer, W. P. C., Rapid evolution of a protein in-vitro by dna shuffling. *Nature* **1994,** *370* (6488), 389-391.
21. Ardejani, M. S.; Li, N. X.; Orner, B. P., Stabilization of a Protein Nanocage through the Plugging of a Protein-Protein Interfacial Water Pocket. *Biochemistry* **2011,** *50* (19), 4029-4037.
22. Kelly, S. M.; Jess, T. J.; Price, N. C., How to study proteins by circular dichroism. *Biochimica Et Biophysica Acta-Proteins and Proteomics* **2005,** *1751* (2), 119-139.

Summary, final conclusions, and future directions

Protein nanocages can act as model systems to understand the roles of protein folding, protein-protein interactions and symmetry, in assembling protein quaternary structures and nanoarchitectures. Understanding these fundamentals can help realise the numerous applications that have been proposed to exploit the unique properties of protein nanocages among which are their nanoscale, hollow architecture, water solubility, self-assembly and genomic encodability. This thesis chronicles a body of work which explored the fundamentals of protein nanocage structure (Chapter 3), made inroads toward directing protein nanocages toward nanotechnological applications (Chapter 2) and developed technologies to enable both applied and fundamental protein nanocage research in the future (Chapters 4, 5 and 6).

Chapter 2 explored methods to apply the ferritin nanocages Bfr, Dps and HsFn to the burgeoning field of nanotechnology. The first half of the chapter described attempts to expand our technique to produce protein nanocage encapsulated nanoparticles. Our laboratory had previously devised a double reduction technique to generate gold nanoparticles inside HsFn. The goal of the research described in the first half of Chapter 2 was to apply this technique to Bfr and Dps. Although, protein aggregation thwarted direct application of the original procedure, optimisation was performed with moderate success. In the future, more work needs to be done to further optimise the conditions for these proteins and to determine if lessons learned through this exercise can be carried through to other protein nanocages. In general, the challenges experienced in this project underlined how even application-based science could be aided by fundamental research. They also emphasized the potential utility of developing a methodology for screening multiple particle generation conditions in a rapid manner for their tendency to destabilise protein nanocages. A methodology that does this was achieved in Chapter 5. The second half of Chapter 2 set out to establish initial protocols and controls for the eventual application of protein nanocages to delivery and imaging. If protein nanocages are to be used for transport of cellular cargo to specific tissues and cell types, it is

important to establish if they have any background targeting preference when not ligated to a tissue- or cell-type specific ligand. Therefore we labelled HsFn, which was demonstrated to be a uniquely stable ferritin in the first half of Chapter 2, with ^{111}In and probed its localisation in mice with SPECT/CT. The imaging was highly successful, establishing a baseline for future specific targeting and perhaps suggesting some degree of innate specificity to dorsal lymph nodes and spleen.

The research described in Chapter 3 answered fundamental questions concerning the self-assembly of protein nanocages. Previously our laboratory demonstrated that if a C-terminal helical domain (the E-helix) from Bfr, a maxi-ferritin, were fused to the mini-ferritin Dps, a 12-meric protein, Dps+E, a size intermediate between Bfr and Dps was generated. Chapter 3 describes a crystallography project to determine if the E-domain is flipped inside or outside the nanocage and if the resulting protein adopted the tetrahedral symmetry of Dps or the octahedral symmetry of Bfr. We successfully obtained the crystal structure and determined that Dps+E is tetrahedral and the E-domain is flipped out of the nanocage. Unfortunately, aside from a few residues, the electron density along the E-domain is weak, preventing us from fully characterising its structure. A future direction of the project is to perform high resolution electron microscopy and determine the structure of Dps+E with the help of tomographic analysis of the micrographs. In addition, we think that the E-domain projected from the nanocage could have future applications as a handle, either for appending directing ligands for delivery applications or to help supra-assemble protein-based materials.

Chapters 4, 5 and 6 describe the development of a new tool which should enable both fundamental research into protein nanocages and protein nanocage applications. With the challenges associated with protein aggregation during the nanoparticle studies described in Chapter 2 and the general frustration our laboratory has experienced with the slow progress of protein nanocage characterisation, it was realised that an assay to directly detect protein

nanocage oligomerization was needed. This method needed to be rapid, and we hoped it would be applicable to complex solutions like lysates so as to bypass time- and resource-intensive protein purification, and we needed it to be readily expandable to high throughput so that we could screen multiple *in vitro* conditions such as those for conjugation of SPECT labels or for the generation of nanoparticles. We also wanted a technique that could work in live cells so that we could understand the role of protein nanocage assembly on cell or virus biology or for the screening of protein libraries.

Chapter 4 describes the development of a bipartite strategy to design, across oligomerization dependent protein-protein interfaces, binding sites on Bfr for the fluorescent imagining reagent FlAsH. Control experiments eventually demonstrated that FlAsH binding was not specific for the nanocage oligomerization state. It is possible the *in vitro* conditions used could be optimized in the future; the experience that was gained by working with this strategy in Bfr was ported to Dps with the hope that the mini-ferritin would be more amenable to the technique.

In Chapter 5 we demonstrated that the mini-ferritin Dps worked exactly as designed and we were able to show that the FlAsH fluorphore and bipartite tetracysteine display technology could be applied to large nanocage proteins to create a direct assay that is specific for the nanocage oligomerization state. The method was readily expanded to an *in vitro,* medium throughput assay to screen potential conditions for gold nanoparticle generation. In the future, this data and similar experiments could be used to overcome challenges in materials applications related to those described in the first half of Chapter 2. It is also thought that this *in vitro* technique could be developed to identify assembly intermediates and to possibly characterise the kinetics of their formation and consumption, thus shedding light on the fundamental mechanism of protein nanocage oligomerization.

Chapter 6 follows directly on the strong success of the research in Chapter 5. We realised that if we could port the FlAsH assay into living bacteria we could perform high throughput screens of protein libraries using cell sorting to discover proteins with enhanced nanocage forming properties. Using microscopy and flow cytometry, we demonstrated that an *in vivo* assay was possible. As a proof of principle library screen we randomised the protein-protein interface of a Dps mutant that was crippled in its ability to assemble into a nanocage. After four rounds of screening with cell sorting, we isolated library members with recovered nanocage formation emphasizing the success of the strategy. Like any enabling technology, the methods described in Chapters 5 and 6 could go in many future directions. Optimised library generation conditions are clearly required, but the technique itself could be used to further optimise bipartite FlAsH binding site linkers, with further expansion into other nanocages. Along with applying this approach to other protein nanocages, we are also interested in screening for function; perhaps conditional assembly could be achieved by developing a "chemical inducers of oligomerization" strategy using small molecules that are photoisomerisable. Other interfaces of the Dps protein could also be screened to enhance stability and interface packing to overcome the crippled assembly mutations.

Taken together, the research projects presented in this thesis have explored the fundamentals of protein nanocage assembly, have made inroads into protein nanocage-based nanomaterials applications, and have developed an enabling technology which should support both fundamental and application based protein nanocage research. The work presented here has added to scientific understanding of this ubiquitous class of proteins and will help protein nanocages fulfil their potential in multiple scientific fields.

Appendix

Appendix contents

Appendix

Appendix for Chapter 2

Dps

MSTAKLVKSKATNLLYTRNDVSDSEKKATVELLNRQVIQFIDLSLITKQAHWNMRGANFIAVHEMLDGFRT
ALIDHLDTMAERAVQLGGVALGTTQVINSKTPLKSYPLDIHNVQDHLKELADRYAIVANDVRKAIGEAKDDD
TADILTAASRDLDKFLWFIECNIE

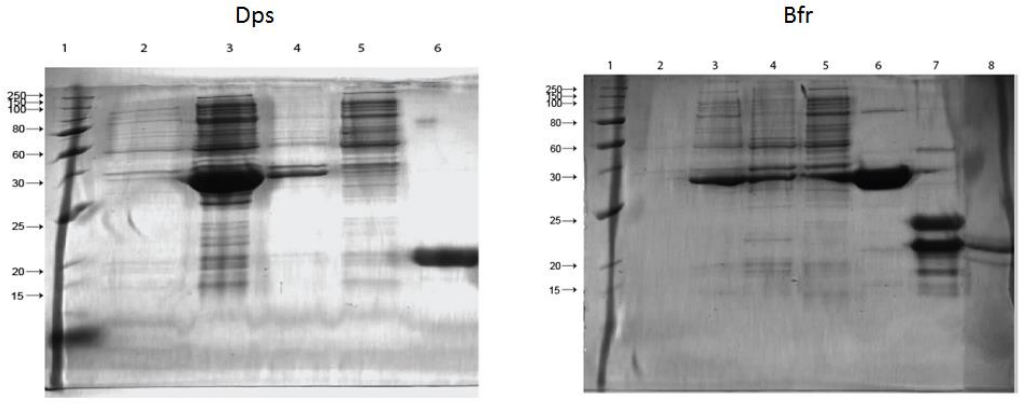
HsFn

MSSQIRQNYSTEVEAAVNRLVNLYLRASYTYLSLGFYFDRDDVALEGVCHFFRELAEEKREGAERLLKMQNQ
RGGRALFQDLQKPSQDEWGTTLDAMKAAIVLEKSLNQALLDLHALGSAQADPHLCDFLESHFLDEEVKLIKK
MGDHLTNIQRLVGSQAGLGEYLFERLTLKHD

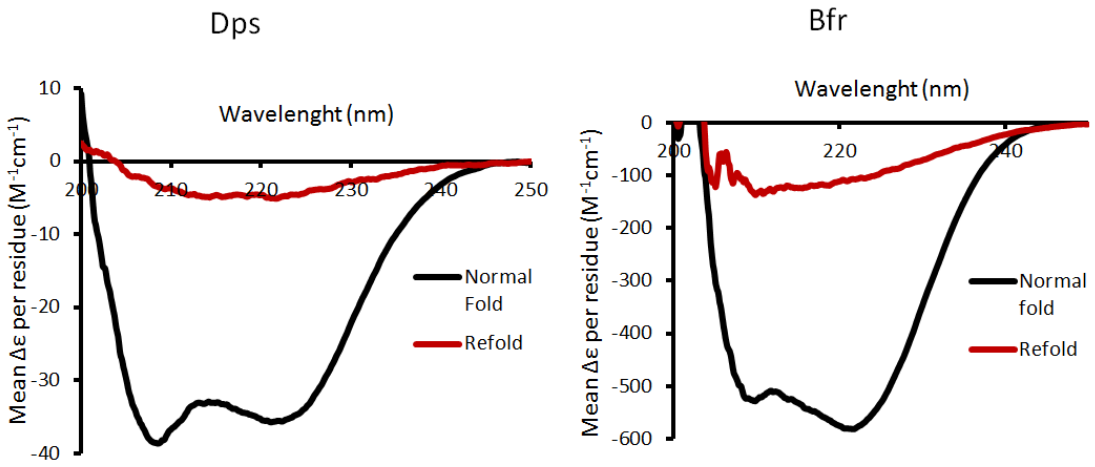
Bfr

MKGDTKVINYLNKLLGNELVAINQYFLHARMFKNWGLKRLNDVEYHESIDEMKHADRYIERILFLEGLPNLQ
DLGKLNIGEDVEEMLRSDLALELDGAKNLREAIGYADSVHDYVSRDMMIEILRDEEGHIDWLETELDLIQKM
GLQNYLQAQIREEG

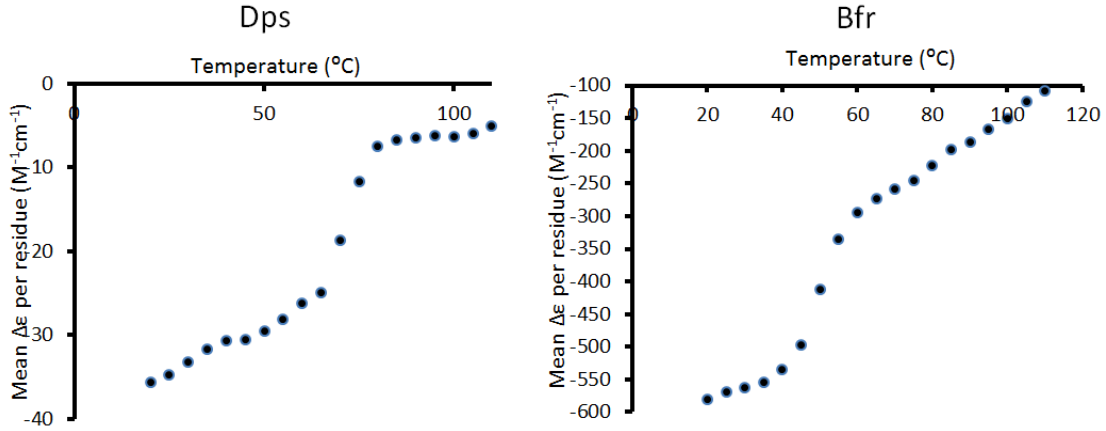
Appendix 2.1 – Amino acid (single letter code) sequences for the wild type monomers used in Chapter 2, DNA binding protein from starved cells (Dps)[1], Horse spleen light chain ferritin (HsFn)[2] and Bacterioferritin (Bfr)[3].



Appendix 2.2 - SDS-PAGE gel (15%) analysis for (left) Dps and (right) Bfr during purification. (both gels) Lane 1: Protein standard ladder (NEB), Lane 2: insoluble fraction, lane 3: soluble fraction, lane 4: wash from first histrap, lane 5: elution from first histrap. (left gel) lane 6: SEC after Ek digestion and second histrap. (right gel) Lane 6: elution from second histrap after Ek digestion, lane 7: wash from second histrap after EK digestion, lane 8: SEC.



Appendix 2.3 - Circular Dichroism spectra demonstrating the effect of heating on the fold of (left) Dps and (right) Bfr. Black: CD signal before heating denaturation. Red: CD signal after cooling post heating denaturation. Protein set to 0.2 mg/ml in phosphate buffer (50 mM Na_2HPO_4, 50 mM NaCl, pH 7) performed on JASCO J-810 spectropolarimeter from 250 nm to 200 nm with a path length of 0.5 mm.



Appendix 2.4 - Thermal melting plots of wild type Dps and Bfr at 222 nm in CD thermal melts were performed using 0.2 mg/ml protein in phosphate buffer, 50 mM Na_2HPO_4, 50 mM NaCl, pH 7 in a range of 4 to 100 °C with a path length of 0.5 mm on a JASCO J-810 spectropolarimeter from 250 nm to 200 nm. Data fitted to a two state unfolding (m1+(m2-m1)/(1+(m0/m3)^m4)equation[4, 5, 6] which calculated the melting temperature of Dps – 68.2 ^{0}C and Bfr – 49.5 ^{0}C. Data fitted using Kaleidagraph(Synergy).

0.1M $NaBH_4$

		Reaction buffer - Phosphate pH 7			Reaction Buffer - Tris pH 7.8			Reaction Buffer - dH_2O		
		Buffer for $NaBH_4$			Buffer for $NaBH_4$			Buffer for $NaBH_4$		
		Na_2HPO_4	Tris	dH_2O	Na_2HPO_4	Tris	dH_2O	Na_2HPO_4	Tris	dH_2O
Buffer for $HAuCl_4$	Na_2HPO_4	Purple	Purple	Purple	Purple	Purple	Purple	Purple	Purple	Purple
	Tris	Purple	Purple	Purple	Purple	Red	Red	Purple	Red	Red
	dH_2O	Purple	Purple	Purple	Purple	Red	Purple	Purple	Purple	Purple

0.01M $NaBH_4$

		Reaction buffer - Phosphate pH 7			Reaction Buffer - Tris pH 7.8			Reaction Buffer - dH_2O		
		Buffer for $NaBH_4$			Buffer for $NaBH_4$			Buffer for $NaBH_4$		
		Na_2HPO_4	Tris	dH_2O	Na_2HPO_4	Tris	dH_2O	Na_2HPO_4	Tris	dH_2O
Buffer for $HAuCl_4$	Na_2HPO_4	Purple	Purple	Purple	Purple	Purple	Purple	Purple	Purple	Purple
	Tris	Purple	Purple	Purple	Purple	Red	Red	Purple	Red	Red
	dH_2O	Purple	Purple	Purple	Purple	Red	Red	Purple	Purple	Purple

0.001M $NaBH_4$

		Reaction buffer - Phosphate pH 7			Reaction Buffer - Tris pH 7.8			Reaction Buffer - dH_2O		
		Buffer for $NaBH_4$			Buffer for $NaBH_4$			Buffer for $NaBH_4$		
		Na_2HPO_4	Tris	dH_2O	Na_2HPO_4	Tris	dH_2O	Na_2HPO_4	Tris	dH_2O
Buffer for $HAuCl_4$	Na_2HPO_4	Purple	Purple	Purple	Purple	Purple	Purple	Purple	Purple	Purple
	Tris	Yellow	Yellow	Yellow	Red	Red	Red	Red	Red	Red
	dH_2O	Purple	Purple	Purple	Purple	Red	Red	Purple	Purple	Purple

Appendix 2.5 – Colourmetric screen to evaluate the ability of $NaBH_4$ at different concentrations to reduce gold in different buffering solutions by observing colour change with red/purple being seen as $NaBH_4$ maintaining its ability to reduce gold. Reaction buffer is 1 ml of solution either Phosphate buffer (50 mM Na_2HPO_4, pH 7), Tris buffer (50 mM Tris.HCl, pH 7.8) or unbuffered in water, to which 20 µl of 0.1 M $HAuCl_4$ in either a solution of Tris, phosphate or water, followed by 20 µl of 0.1 M $NaBH_4$ also in either a solution of Tris, phosphate or water and left to incubate at room temperature for 12 hours.

0.1M $NaCNBH_3$

		Reaction buffer - Phosphate pH 7			Reaction Buffer - Tris pH 7.8			Reaction Buffer - dH_2O		
		Buffer for $NaCNBH_3$			Buffer for $NaCNBH_3$			Buffer for $NaCNBH_3$		
		Na_2HPO_4	Tris	dH_2O	Na_2HPO_4	Tris	dH_2O	Na_2HPO_4	Tris	dH_2O
Buffer for $HAuCl_4$	Na_2HPO_4	Purple	Purple	Purple	Purple	Purple	Purple	Purple	Purple	Purple
	Tris	Purple	Purple	Purple	Purple	Purple	Purple	Purple	Purple	Purple
	dH_2O	Purple	Purple	Purple	Purple	Purple	Purple	Purple	Purple	Purple

0.01M $NaCNBH_3$

		Reaction buffer - Phosphate pH 7			Reaction Buffer - Tris pH 7.8			Reaction Buffer - dH_2O		
		Buffer for $NaCNBH_3$			Buffer for $NaCNBH_3$			Buffer for $NaCNBH_3$		
		Na_2HPO_4	Tris	dH_2O	Na_2HPO_4	Tris	dH_2O	Na_2HPO_4	Tris	dH_2O
Buffer for $HAuCl_4$	Na_2HPO_4	No change	Purple	Purple	No change	Purple	Purple	No change	Purple	Purple
	Tris	Yellow	Purple	Purple	Yellow	Purple	Purple	Yellow	Purple	Red
	dH_2O	No change	Purple	Purple	No change	Purple	Purple	No change	Purple	Red

0.001M $NaCNBH_3$

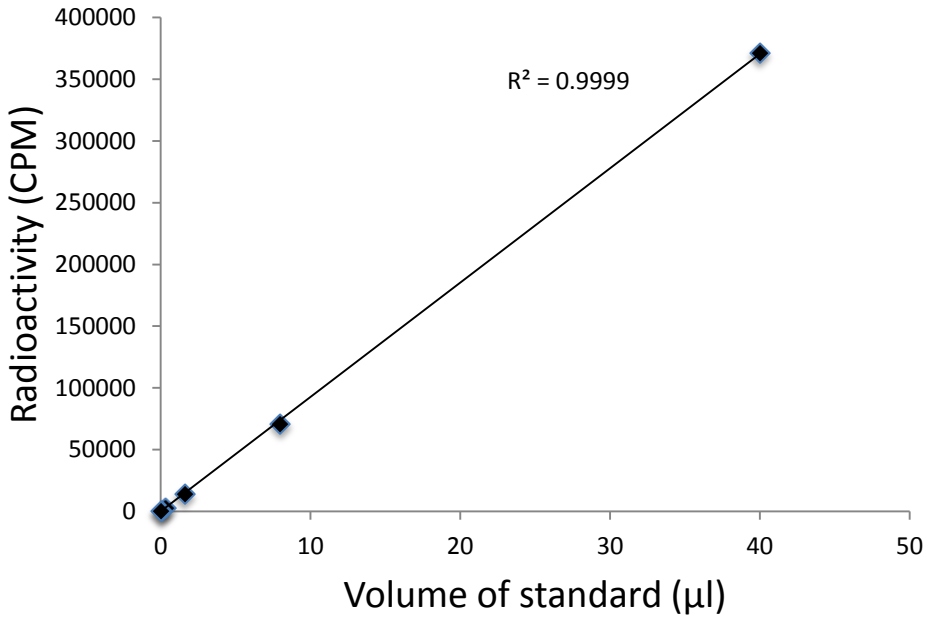
		Reaction buffer - Phosphate pH 7			Reaction Buffer - Tris pH 7.8			Reaction Buffer - dH_2O		
		Buffer for $NaCNBH_3$			Buffer for $NaCNBH_3$			Buffer for $NaCNBH_3$		
		Na_2HPO_4	Tris	dH_2O	Na_2HPO_4	Tris	dH_2O	Na_2HPO_4	Tris	dH_2O
Buffer for $HAuCl_4$	Na_2HPO_4	No change	No change	No change	No change	No change	Purple	No change	No change	Purple
	Tris	Yellow	Yellow	Yellow	Yellow	Yellow	Purple	Yellow	Yellow	Red
	dH_2O	No change	No change	No change	No change	No change	Purple	No change	No change	Purple

Appendix 2.6 - Screen to evaluate the ability of $NaCNBH_3$ at different concentrations to reduce gold in different buffering solutions by observing colour change with red/purple being seen as $NaCNBH_3$ maintaining its ability to reduce gold. Reaction buffer is 1 ml of solution either Phosphate buffer (50 mM Na_2HPO_4, pH 7), Tris buffer (50 mM Tris.HCl, pH 7.8) or unbuffered in water, to which 20 µl of 0.1 M $HAuCl_4$ in either a solution of Tris, phosphate or water, followed by 20 µl of 0.1 M $NaCNBH_3$ also in either a solution of Tris, phosphate or water and left to incubate at room temperature for 12 hours**.**

Trial	First gold addition			First reducing agent			Second gold addition			Second reducing agent			Other	Bfr Result	Dps Result
	Volume (µl)	Concentration (M)	Buffer	Volume (µl)	Concentration (M)	Buffer	Volume (µl)	Concentration (M)	Buffer	Volume (µl)	Concentration (M)	Buffer			
	$HAuCl_4$			$NaBH_4$			$HAuCl_4$			Ascorbic acid					
From paper	22.6	0.1	H_2O	20	0.1	H_2O	10	0.1	H_2O	30	0.1	H_2O		1	1
1	22.6	0.1	H_2O	20	0.1	H_2O	10	0.1	H_2O	30	0.1	H_2O	Incubation at 4 °C	1	1
2	22.6	0.1	H_2O	20	0.1	H_2O	10	0.1	H_2O	30	0.1	H_2O	Addition of 0.1% Tween20	1	1
3	22.6	0.1	H_2O	20	0.01	H_2O	10	0.1	H_2O	30	0.1	H_2O		1	1
4	22.6	0.1	H_2O	20	0.001	H_2O	10	0.1	H_2O	30	0.1	H_2O		1	1
5	50	0.05	H_2O	20	0.1	H_2O	10	0.1	H_2O	30	0.1	H_2O		1	1
6	22.6	0.1	Tris	20	0.1	H_2O	10	0.1	Tris	30	0.1	H_2O		1	2
7	50	0.05	Tris	20	0.1	H_2O	20	0.05	Tris	30	0.1	H_2O		1	2
8	22.6	0.1	Tris										Examined after 3 hours	2	2
9	50	0.05	Tris										Examined after 3 hours	3	3
10				20	0.1	Tris							Examined after 3 hours	1	2
11				20	0.01	Tris							Examined after 3 hours	1	3
	$HAuCl_4$			$NaCNBH_3$			$HAuCl_4$			Ascorbic acid					
12				20	0.1	Tris							Examined after 3 hours	3	3
13				20	0.01	Tris							Examined after 3 hours	3	3
14				100	0.1	Tris							Examined after 3 hours	3	3
15				200	0.1	Tris							Examined after 3 hours	3	3
16	20	0.1	Tris	20	0.1	Tris	10	0.1	Tris	30	0.1	H_2O		2	2
17	5x20	0.05	Tris	20	0.1	Tris	5x10	0.05	Tris	30	0.1	H_2O		3	3
18	100	0.05	Tris	20	0.1	Tris	100	0.05	Tris	30	0.1	H_2O		3	3
19	20	0.1	Tris	100	0.1	Tris	10	0.1	Tris	30	0.1	H_2O		2	3
20	5x20	0.05	Tris	100	0.1	Tris	5x10	0.05	Tris	30	0.1	H_2O		3	3

21	100	0.05	Tris	100	0.1	Tris	100	0.05	Tris	30	0.1	H_2O		3	3
22	5x20	0.05	Tris	100	0.1	Tris	10	0.1	Tris	50	0.1	Tris		4 #	3 #
23	5x40	0.05	Tris	100	0.1	Tris	10	0.1	Tris	50	0.1	Tris		4 #	3 #
24	100	0.05	Tris	100	0.1	Tris	10	0.1	Tris	50	0.1	Tris		4 #	3 #
25	5x20	0.05	Tris	100	0.1	Tris	100	0.05	Tris	50	0.1	Tris		4 #	4 #
26	5x40	0.05	Tris	100	0.1	Tris	100	0.05	Tris	50	0.1	Tris		4 #	4 #
27	100	0.05	Tris	100	0.1	Tris	100	0.05	Tris	50	0.1	Tris		4 #	4 #

Appendix 2.7 – Summary of the optimisation conditions employed for the genesis of gold nanoparticles inside of the ferritins Dps and Bfr. This table lists the volume added, concentration and buffering solution either Tris (50 mM Tris.HCl, pH 7.8) or unbuffered in water, of the first gold addition, the first reducing agent, the second gold addition and second reducing agent. The results of each trial are recorded by number. 1 – Condition lead to total aggregation of the protein. 2 – Condition lead to almost complete protein aggregation but with a much longer elution UV 280 nm signal with low intensity but with no clear nanocage peak. 3 – Condition lead to slight aggregation, with a longer elution UV 280 nm signal with higher intensity but with no clear nanocage peak. 4 – Condition lead to the presence of a small protein aggregate peak with a longer UV 280 nm elution signal of high intensity but with a well-defined nanocage peak. # - Indicates the presence of a SPR peak with a defined peak corresponding to the elution volume of the nanocage proteins.



Appendix 2.8 – Radioactivity from serial dilutions of the standard radiolabelled conjugated HsFn sample containing 0.1 mg of protein, 25 µl of ^{111}In with 193 Mbq of activity which was diluted to 100 µl with saline before undergoing serial dilutions. Used as a comparison to monitor any uptake into the mice.

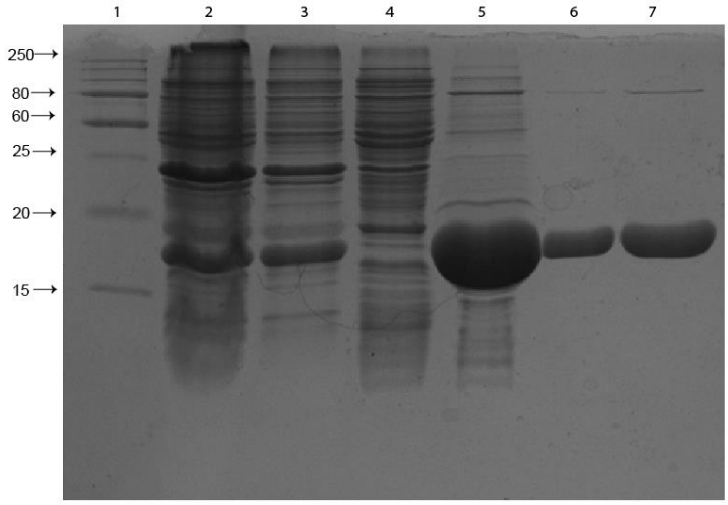
Appendix for Chapter 3

Primer name	Primer sequences (5'-3')
Dps+E F	GACGACGACAAGATGAGCACCGCAAAACTGGTG
Dps+E R	GAGGAGAAGCCCGGTTCAACCTTCTTCACGGAT

Appendix 3.1 – Primer used to transfer the gene for the protein Dps+E from pET-32b into pET-46. Primers sourced from Suprenom Singapore, F – forward primer, R – Reverse primer.

MAHHHHHHVDDDDKMSTAKLVKSKATNLLYTRNDVSDSEKKATVELLNRQVIQFIDLSLITKQAHWNMR
GANFIAVHEMLDGFRTALIDHLDTMAERAVQLGGVALGTTQVINSKTPLKSYPLDIHNVQDHLKELADRYAI
VANDVRKAIGEAKDDDTADILTAASRDLDKFLWFIECNLDLIQKMGLQNYLQAQIREEG

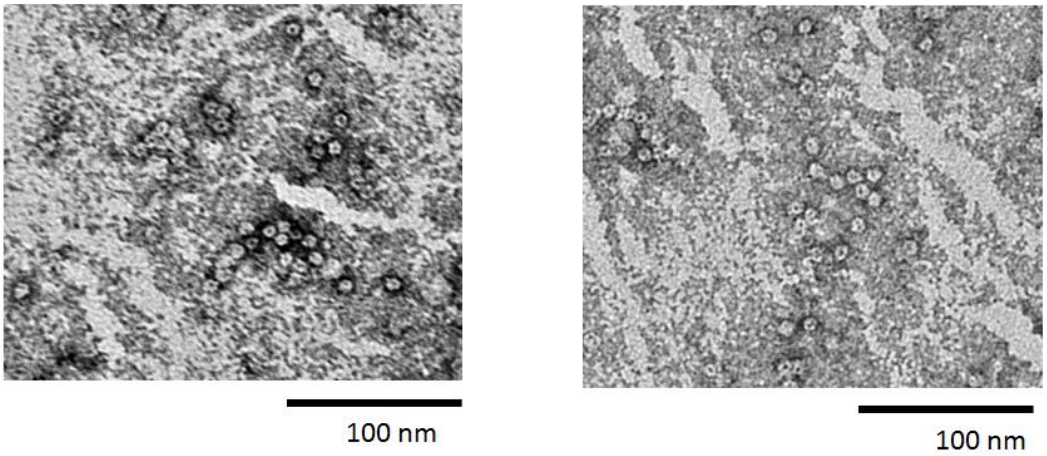
Appendix 3.2 - Amino acid (one letter code) sequences for the hybrid ferritin Dps+E monomers used in Chapter 3[4].



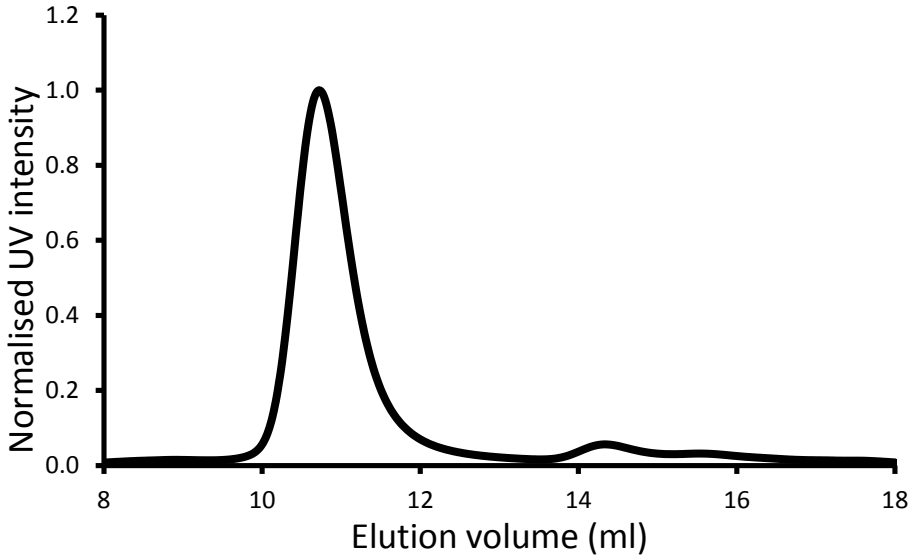
Appendix 3.3 - SDS-PAGE gel (15%) analysis for the purification of Dps+E. Lane 1: Protein standard ladder (NEB), Lane 2: soluble fraction, lane 3: insoluble fraction, lane 4: wash from first histrap, lane 5: elution from first histrap, lane 6: IEX, lane 7: SEC.

Protein	Measured mass (Da)	Expected mass (Da)	Difference (Da)
Dps+E	22460	22493.4	33.4

Appendix 3.4 – Maldi-TOF mass spectrometry analysis of the purified protein Dps+E. Expected mass is minus the N-terminal methionine.



Appendix 3.5 – Transmission electron microscopy micrographs to show that Dps+E can form a nanocage, using 1% uranyl acetate as a negative stain. Images made using ImageJ[7].



Appendix 3.6 - Size exclusion chromatogram of purified Dps+E. Chromatogram normalised to its highest intensity and the average of three runs (0.5 ml injection of 1 mg/ml protein (followed at 280 nm), using running buffer, 50 mM Na_2HPO_4, 50 mM NaCl, pH 7 at a flow rate of 0.5 ml/min.

1

	5%	10%	15%	20%	25%
PEG 400					
PEG 3350					

PEG 6000					

All cells contain 0.1 M Tris pH 8.6

2

PEG 400

	5%	10%	15%	20%	25%
0.2 M Lithium sulfate					
0.2 M Magnesium chloride					
0.2 M Tri Sodium citrate					

All cells contain 0.1 M Tris pH 8.6

3

PEG 3350

	5%	10%	15%	20%	25%
0.2 M Lithium sulfate					
0.2 M Magnesium chloride					
0.2 M Tri Sodium citrate					

All cells contain 0.1 M Tris pH 8.6

4

PEG 6000

	5%	10%	15%	20%	25%
0.2 M Lithium sulfate					
0.2 M Magnesium chloride					
0.2 M Tri Sodium citrate					

All cells contain 0.1 M Tris pH 8.6

5

PEG 400

	30%	35%	40%	45%	50%
0.2 M Lithium sulfate					
0.2 M Magnesium chloride					
0.2 M Tri Sodium citrate					

All cells contain 0.1 M Tris pH 8.6

6

PEG 3350

	12%	14%	16%	18%	20%
0.1 M Tris pH 8.4					
0.1 M Tris pH 8.6					
0.1 M Tris pH 8.8					

All cells contain 0.2 M Magnesium chloride

7

0.2 M Magnesium Formate	10% PEG 3350	15% PEG 3350	20% PEG 3350	25% PEG 3350	30% PEG 3350
0.2 M Lithium sulfate, 0.1M sodium acetate pH 4.2	10% PEG 1000	15% PEG 1000	20% PEG 1000	25% PEG 1000	30% PEG 1000
0.1 M sodium acetate pH 4.6	20% PEG 400	25% PEG 400	30% PEG 400	35% PEG 400	40% PEG 400

8

PEG 6000

	12%	14%	16%	18%	20%
0.1 M Tris pH 8.2					
0.1 M Tris pH 8.6					
0.1 M Tris pH 8.8					

All cells contain 0.2 M Magnesium chloride

9

PEG 6000

	12%	14%	16%	18%	20%
0.1 M Tris pH 8.2					
0.1 M Tris pH 8.6					
0.1 M Tris pH 8.8					

All cells contain 0.2 M Lithium sulfate

10

PEG 3350

	12%	14%	16%	18%	20%
0.1 M Magnesium formate					
0.2 M Magnesium formate					
0.3 M Magnesium formate					

All cells contain 0.1 M Tris pH 8.6

11

PEG 1000

	16%	18%	20%	22%	24%
0.1 M Sodium acetate trihydrate pH 4					
0.1 M Sodium acetate trihydrate pH 4.2					
0.1 M Sodium acetate trihydrate pH 4.4					

All cells contain 0.2 M Lithium sulfate

12

PEG 3350

	14%	15%	16%	17%	18%
0.1 M Magnesium chloride					
0.2 M Magnesium chloride					
0.3 M Magnesium chloride					

All cells contain 0.1 M Tris pH 8.6

13

PEG 3350

	14%	15%	16%	17%	18%
0.1 M Tris pH 8					
0.1 M Tris pH 8.6					
0.1 M Tris pH 9					

All cells contain 0.2 M Magnesium chloride

PEG 3350

14

	0.2 M $MgCl_2$	0.25 M $MgCl_2$	0.3 M $MgCl_2$	0.35 M $MgCl_2$	0.4 M $MgCl_2$
14% PEG 3350					
15% PEG 3350					
16% PEG 3350					

All cells contain 0.1 M Tris pH 8.6

15

	0.2 M $C_2H_2O_4Mg$	0.25 M $C_2H_2O_4Mg$	0.3 M $C_2H_2O_4Mg$	0.35 M $C_2H_2O_4Mg$	0.4 M $C_2H_2O_4Mg$
14% PEG 3350					
15% PEG 3350					
16% PEG 3350					

PEG 3350

16

	12%	12.5%	13%	13.5%	14%
0.3 M Magnesium chloride					
0.4 M Magnesium chloride					
0.5 M Magnesium chloride					

All cells contain 0.1 M Tris pH 8.6

PEG 400

17

	20%	21%	22%	23%	24%
0.1 M Sodium acetate trihydrate pH 4					
0.1 M Sodium acetate trihydrate pH 4.6					
0.1 M Sodium citrate tribasic dihydrate pH 5					

PEG 3350

18

	16%	18%	20%	22%	24%
0.1 M Tri sodium citrate					
0.2 M Tri sodium citrate					
0.3 M Tri sodium citrate					

All cells contain 0.1 M Tris pH 8.6

PEG 8000

19

	8%	9%	10%	11%	12%
0.1 M Calcium acetate					
0.2 M Calcium acetate					
0.3 M Calcium acetate					

All cells contain 0.1 M HEPES pH 7.4

PEG 400

20

	17%	18%	19%	20%	21%
0.1 M Sodium citrate tribasic dihydrate pH 5					
0.1 M Sodium citrate tribasic dihydrate pH 5.6					
0.1 M Sodium cacodylate trihydrate pH 6					

PEG 3350

6

	18%	19%	20%	21%	22%
0.3 M Trisodium acetate					
0.2 M Trisodium acetate					
0.3 M Trisodium acetate					

All cells contain 0.1 M Tris pH 8.6

PEG 1000

7

	pH 3.6	pH 4	pH 4.4	pH 4.8	pH 5.2
18% PEG 1000					
20% PEG 1000					
22% PEG 1000					

PEG 1000

21

	14%	16%	18%	20%	22%
0.2 M Lithium sulfate					
0.3 M Lithium sulfate					
0.4 M Lithium sulfate					

All cells contain 0.1 M Sodium acetate trihydrate pH 4.2

PEG 1000

22

	16%	17%	18%	19%	20%
0.1 M Lithium sulfate					
0.15 M Lithium sulfate					
0.2 M Lithium sulfate					

All cells contain 0.1 M Sodium acetate trihydrate pH 4

PEG 1000

23

	16%	17%	18%	19%	20%
0.1 M Lithium sulfate					
0.15 M Lithium sulfate					
0.2 M Lithium sulfate					

All cells contain 0.1 M Sodium acetate trihydrate pH 4.8

Appendix 3.7 – Hand-made 15 well plates for optimisation of conditions to crystallise the protein Dps+E. All wells were hanging drop with grey boxes representing the conditions where

crystals were observed. All conditions recorded above or below each table are present in all the wells.

Position	Amino acid	RMSD (Å)	Position	Amino acid	RMSD (Å)	Position	Amino acid	RMSD (Å)
13	N	0.123	66	L	0.254	124	R	0.044
14	L	0.215	67	D	0.481	125	Y	0.116
15	L	0.125	68	G	0.498	126	A	0.106
16	Y	0.245	69	F	0.296	127	I	0.140
17	T	0.139	70	R	0.360	128	V	0.055
18	R	0.319	71	T	0.328	129	A	0.134
19	N	0.240	72	A	0.356	130	N	0.164
20	D	0.229	73	L	0.243	131	D	0.121
21	V	0.146	74	I	0.177	132	V	0.080
22	S	0.267	75	D	0.275	133	R	0.110
23	D	0.251	76	H	0.213	134	K	0.205
24	S	0.167	77	L	0.165	135	A	0.152
25	E	0.223	78	D	0.155	136	I	0.127
26	K	0.186	79	T	0.182	137	G	0.386
27	K	0.216	80	M	0.211	138	E	0.491
28	A	0.173	81	A	0.363	139	A	0.208
29	T	0.138	82	E	0.337	140	K	0.156
30	V	0.116	83	R	0.206	141	D	0.252
31	E	0.172	84	A	0.141	142	D	0.197
32	L	0.110	85	V	0.178	143	D	0.069
33	L	0.086	86	Q	0.180	144	T	0.151
34	N	0.152	87	L	0.133	145	A	0.118
35	R	0.135	88	G	0.055	146	D	0.113
36	Q	0.151	89	G	0.088	147	I	0.125
37	V	0.171	90	V	0.112	148	L	0.088
38	I	0.118	91	A	0.094	149	T	0.156
39	Q	0.080	92	L	0.079	150	A	0.116
40	F	0.096	93	G	0.159	151	A	0.096
41	I	0.112	94	T	0.189	152	S	0.165
42	D	0.070	95	T	0.164	153	R	0.147
43	L	0.109	96	Q	0.248	154	D	0.140
44	S	0.070	97	V	0.200	155	L	0.198
45	L	0.061	98	I	0.180	156	D	0.083
46	I	0.079	99	N	0.083	157	K	0.078
47	T	0.130	100	S	0.376	158	F	0.101
48	K	0.232	101	K	0.164	159	L	0.127
49	Q	0.101	102	T	0.147	160	W	0.058
50	A	0.206	103	P	0.121	161	F	0.070

51	H	0.471	104	L	0.164	162	I	0.072
52	W	0.150	105	K	0.234	163	E	0.079
53	N	0.108	106	S	0.140	164	C	0.193
54	M	0.194	107	Y	0.075	165	N	0.112
55	R	0.132	108	P	0.128			
56	G	0.135	109	L	0.042			
57	A	0.169	110	D	0.136			
58	N	0.067	111	I	0.137			
59	F	0.066	112	H	0.028			
60	I	0.089	113	N	0.115			
61	A	0.149	114	V	0.128			
62	V	0.096	115	Q	0.116			
63	H	0.173	116	D	0.153			
64	E	0.122	122	A	0.080			
65	M	0.093	123	D	0.053			

Figure 3.8 – RMSD calculations from a comparison of all the residues between the wild type Dps protein (PDB:1DPS[1]) and the obtained Dps+E crystal structure. RMSD calculation made using the Matchmaker tool in Chimera[8].

Appendix for Chapter 4

Table 1

Primer name	Primer sequences (5'-3')
BfrWT F	CATCTACAGCATATGAAAGGTGATACCAAAGTGATCAAC
BfrWT R	GAGGAGAACGGTTAACCTTCTTCACGGATCTGTGCCTGCAG
BfrCC-N-term F	CATCTACAGCATATGTGCTGCAAAGGTGATACCAAAGTGATCAAC
BfrCCPGCC-N-term F	CATCTACAGCATATGTGTTGTCCGGGCTGTTGTAAAGGTGATACCAAAGTGATCAAC

Table 2

Primer name	Primer sequences (5'-3')
BfrCC-C-term F	CAGATCCGTGAAGAAGGTTGCTGCTGACTTCTCCTCAACCATGGCG
BfrCC-C-term R	CGCCATGGTTGAGGAGAAGTCAGCAGCAACCTTCTTCACGGATCTG
BfrCCGPCC-C-term F	CCGTGAAGAAGGTTGCTGCGGCCCGTGCTGCTAACATGGCGATATCGG
BfrCCGPCC-C-term R	CCGATATCGCCATGTTAGCAGCACGGGCCGCAGCAACCTTCTTCACGG
BfrExt-C F	CACAGATCCGTGAAGAAGGTTGTTGACGGGCTTCTCCTCAACCATGGCG
BfrExt-C R	CGCCATGGTTGAGGAGAAGCCCGTCAACAACCTTCTTCACGGATCTGTG
BfrExt-G-C F	CACAGATCCGTGAAGAAGGTGGTTGTTGACTTCTCCTCAACCATGGC
BfrExt-G-C R	GCCATGGTTGAGGAGAAGTCAACAACCACCTTCTTCACGGATCTGTG
BfrN148C F	CTGATCCAGAAAATGGGTCTGCAGTGCTACCTGCAGGCACAGATCCGTGAAG
BfrN148C R	CTTCACGGATCTGTGCCTGCAGGTAGCACTGCAGACCCATTTTCTGGATCAG
BfrM147A,N148C F	CTGATCCAGAAAATGGGTCTGGCCTGCTACCTGCAGGCACAGATCCGTGAAG
BfrM147A,N148C R	CTTCACGGATCTGTGCCTGCAGGTAGCAGGCCAGACCCATTTTCTGGATCAG
BfrQ151C F	GGGTCTGCAGAACTACCTGTGCGCACAGATCCGTGAAGAAGG
BfrQ151C R	GGGTCTGCAGAACTACCTGTGCGCACAGATCCGTGAAGAAGG
BfrE157C F	CTGCAGGCACAGATCCGTGAATGTGGTTGAACCGGGCTTCTCCTCAACC
BfrE157C R	GGTTGAGGAGAAGCCCGGTTCAACCACATTCACGGATCTGTGCCTGCAG

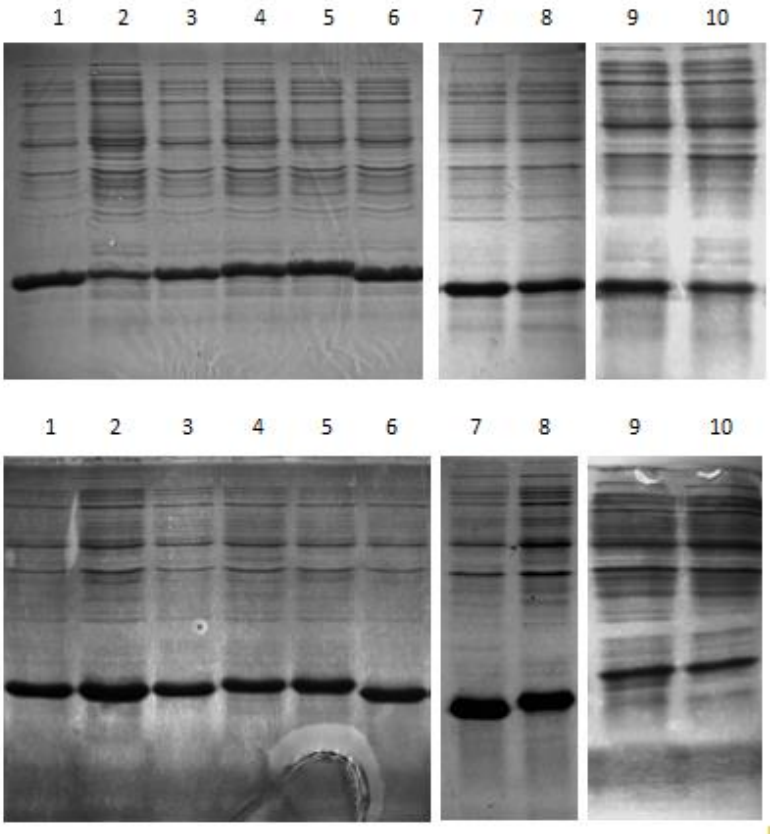
Table 3

Primer name	Primer sequences (5'-3')
BfrN23F F	AACGAACTGGTGGCAATCTTCCAGTACTTCCTGCATGCA
BfrN23F R	TGCATGCAGGAAGTACTGGAAGATTGCCACCAGTTCGTT
BfrD118W F	GTGCATGATTACGTGAGCCGTTTCATGATGATCGAAATCCTGCGT
BfrD118W R	ACGCAGGATTTCGATCATCATGAAACGGCTCACGTAATCATGCAC

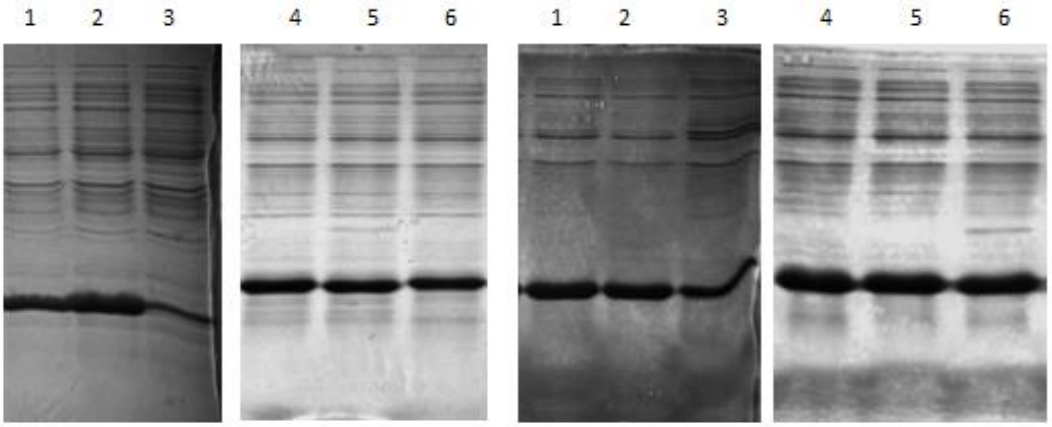
Table 4

Primer name	Primer sequences (5'-3')
BfrWT LIC F	GACGACGACAAGATGAAAGGTGATACCAAAGTGATCAAC
BfrWT LIC R	GAGGAGAACGGTTAACCTTCTTCACGGATCTGTGCCTGCAG
BfrCCPGCC-N-term LIC F	GACGACGACAAGATGTGTTGTCCGGGCTGTTGTAAAGGTGATACCAAAGTGATCAAC
BfrCCPGCC-C-term LIC R	GAGAAGCCCGGTTAGCAGCACGGGCCGCAGCACCATTCTTC
BfrCC-C-term LIC F	GACGACGACAAGATGTGCTGCAAAGGTGATACCAAAGTGATCAAC
BfrCC-C-term LIC R	GAGGAGAAGCCCGGTTAGCAGCAACCTTCTTCACGGATCTGTGCCTGCAG

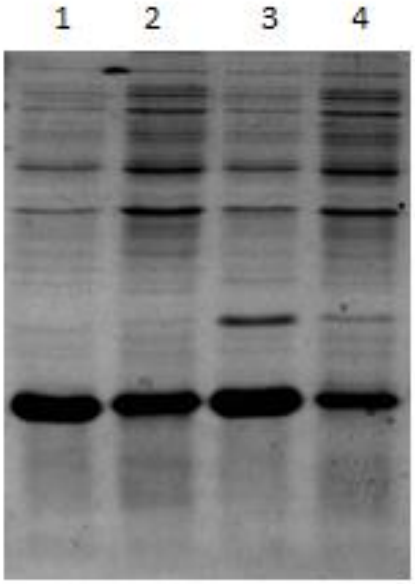
Appendix 4.1 - Primers used for the construction of all mutants and controls used in Chapter 4 for Bfr. Table – Primers used to transfer the wild type Bfr gene from pET-32b to pET-22b and extend the N-terminus. Table 2 – Primers used to construct all pET-22b C-terminus designs and tetrapartite designs. Table 3 – Primers needed to introduce the nanocage stabilising mutations N23F and D118F. Table 4 – Primers needed to transfer successful bipartite designs and control from pET-22b to pEt-32b. All primers were sourced from Suprenom, Singapore. F - forward primer, R – reverse primer**.**



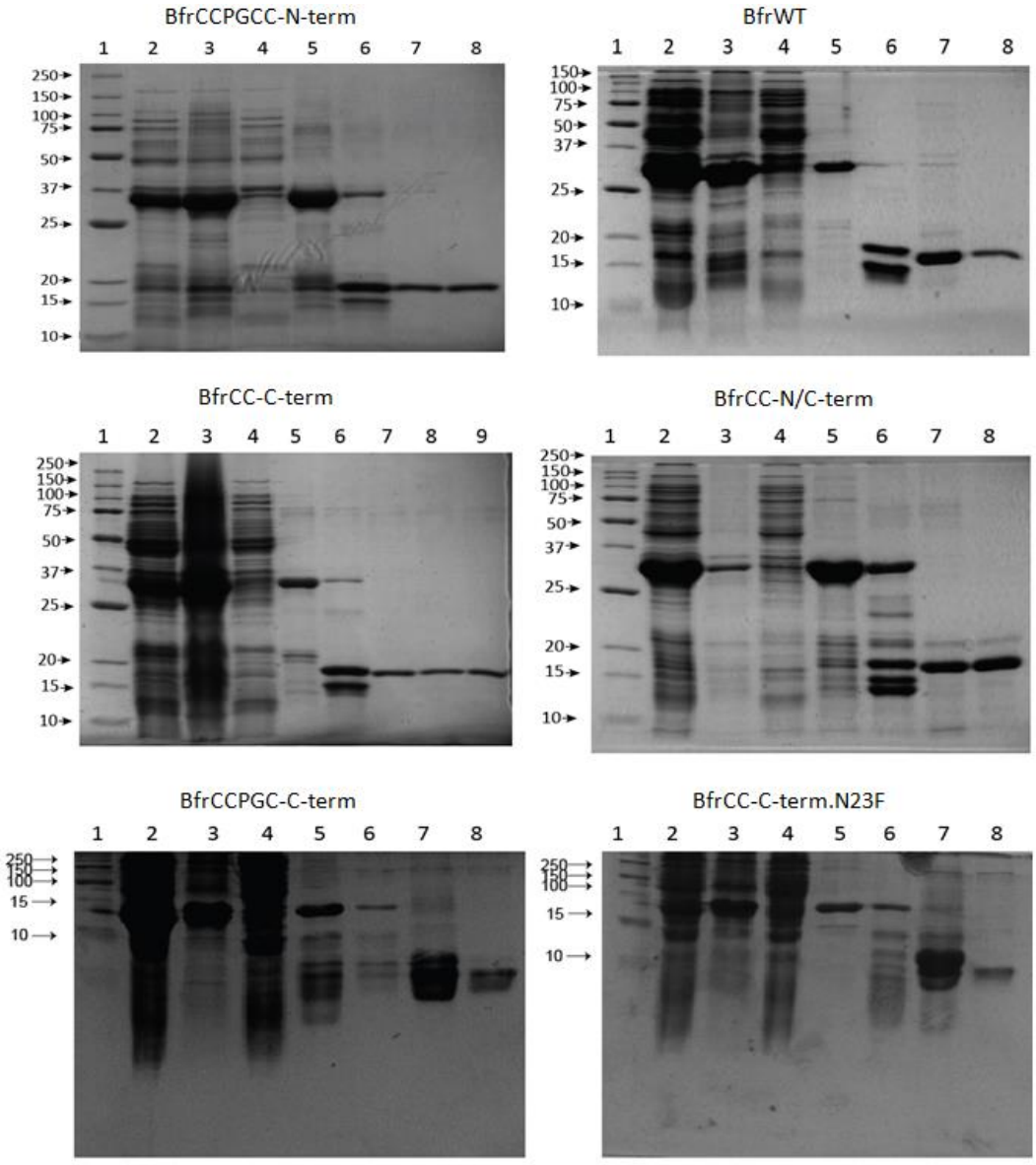
Appendix 4.2 - SDS-PAGE (15%) of lysate samples for each Bfr bipartite designs diluted to 1 mg/ml total protein concentration for comparison of expression levels. Lane 1: **BfrCC-N-term**, lane 2: **BfrCCPGCC-N-term**, lane 3: **BfrCCPGCC-C-term**, lane 4: **BfrCC-C-term**, lane 5: **BfrCC-N/C-term**, lane 6: **BfrWT**, lane 7: **BfrCC-C-termN23f**, lane 8: **BfrCC-N/C-termN23F**, lane 9: **BfrCC-C-termD118F**, and lane 10: **BfrCC-N/C-termD118F**. Each gel (top and bottom) are from different protein expressions of the same designs or controls.

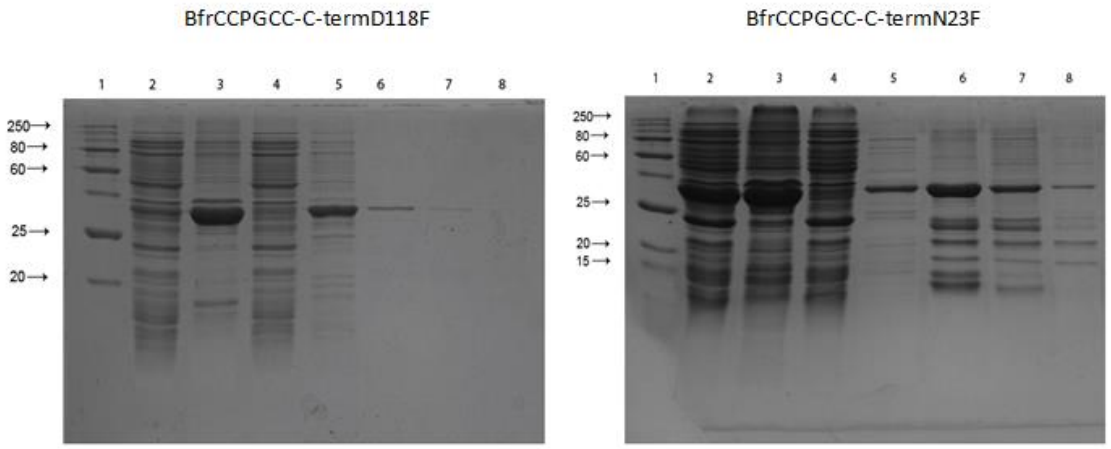


Appendix 4.3 - SDS-PAGE (15%) of lysate samples for each Bfr tetrapartite designs diluted to 1 mg/ml total protein concentration for comparison of expression levels. Lane 1: **BfrN148C**, lane 2: **BfrM147A,N148C**, lane 3: **BfrQ151**C: lane4: **BfrE157C**, lane 5: **BfrExt-C**, and lane 6: **BfrExt-GC**. Each gel (left and right) are from different protein expressions of the same designs or controls.

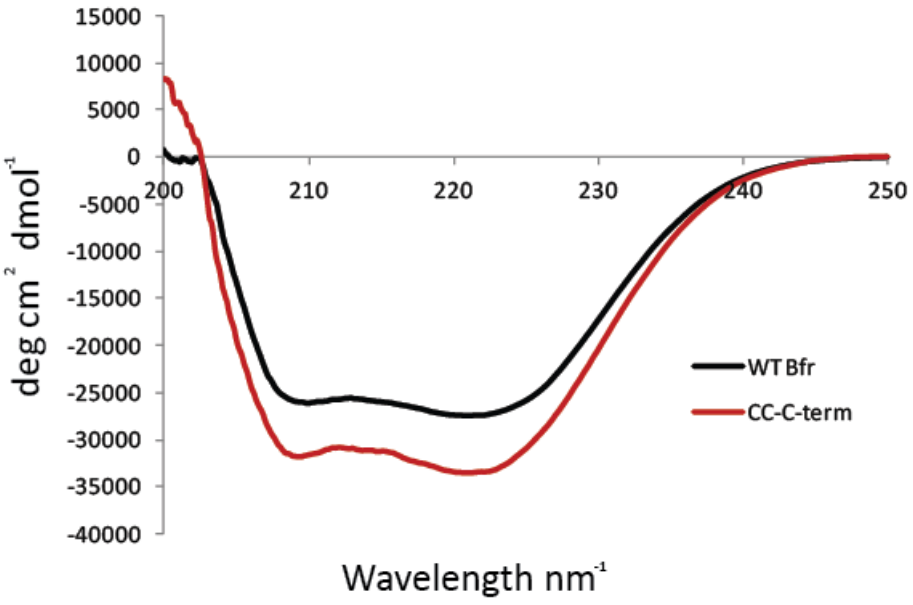


Appendix 4.4 - SDS-PAGE (15%) of lysate samples for each stabilised Bfr tetrapartite designs diluted to 1 mg/ml total protein concentration for comparison of expression levels. Lane 1: **BfrQ151CN23F**, lane 2: **BfrQ151CD118F**, lane 3: **BfrE157CN23F**, and lane 4: **BfrE157CD118F**. Expression analysis via SDS-PAGE performed just once.

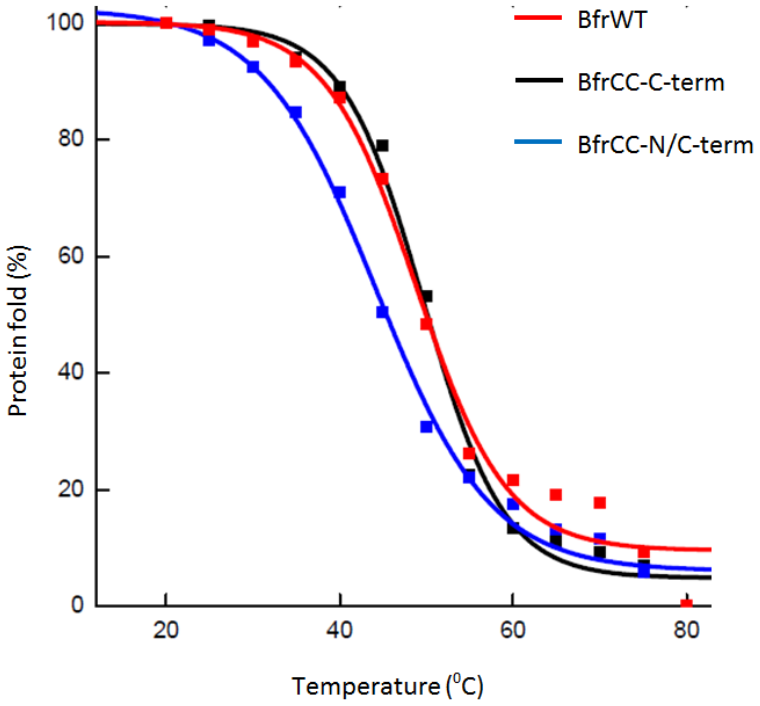

BfrCCPGCC-N-term
1 2 3 4 5 6 7 8
250
150
100
75
50
37
25
20
15
10
BfrWT
1 2 3 4 5 6 7 8
150
100
75
50
37
25
20
15
10
BfrCC-C-term
1 2 3 4 5 6 7 8 9
250
150
100
75
50
37
25
20
15
10
BfrCC-N/C-term
1 2 3 4 5 6 7 8
250
150
100
75
50
37
25
20
15
10
BfrCCPGC-C-term
1 2 3 4 5 6 7 8
250
150
100
15
10
BfrCC-C-term.N23F
1 2 3 4 5 6 7 8
250
150
100
15
10



Appendix 4.5 - SDS-PAGE (15%) analysis of the expression and purification of Bfr designs and controls used in Chapter 4**.** Lane 1: protein ladder 10KDa-250KDa; Lane 2: post induction, soluble fraction; Lane 3: post induction insoluble fraction; Lane 4: wash through from first Histrap; Lane 5: elution from first Histrap; Lane 6: post Enteriokinase digestion, second Histrap elution; Lane7: post Enteriokinase digestion, second Histrap wash; Lane 8: fraction after SEC, cage peak; Lane 9: fraction after SEC, dimer peak.

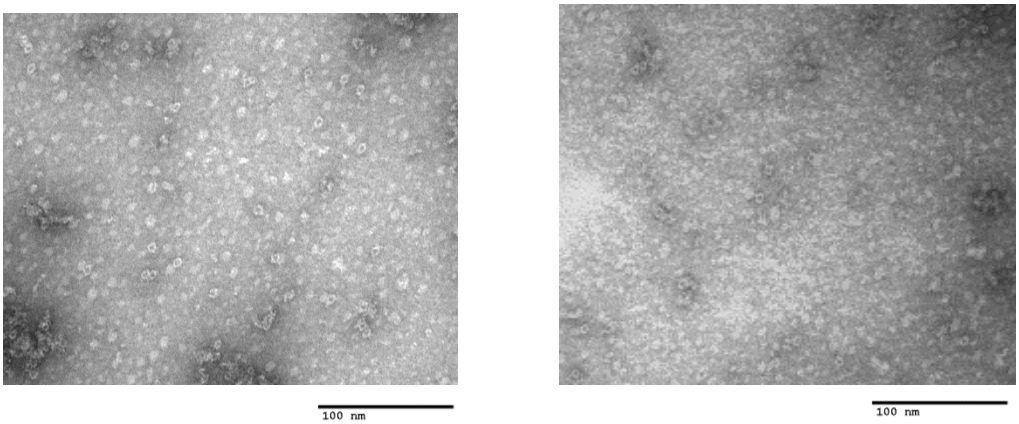


Appendix 4.6 - Circular Dichroism spectra demonstrating that the wild type Bfr protein and the most successful design found in Chapter 4, BfrCC-C-term are folded. Black: CD signal. Protein set to 0.2 mg/ml in phosphate buffer (50 mM Na_2HPO_4, 50 mM NaCl, pH 7) performed on JASCO J-810 spectropolarimeter from 250 nm to 200 nm with a path length of 0.5 mm.

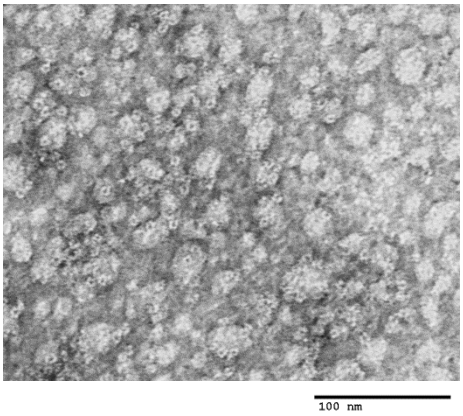


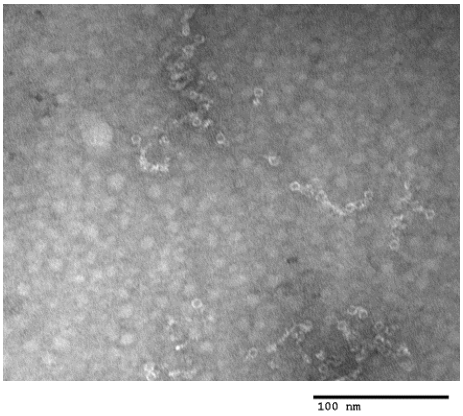
Appendix 4.7 - Thermal melting plots of the Bfr wild type (BfrWT, Red), the purified most interesting bipartite designs (BfrCC-C-term, black and BfrCC-N/C-term, blue) monitored at 222 nm with CD. Thermal melts were performed using 0.2 mg/ml protein in phosphate buffer, 50 mM Na_2HPO_4, 50 mM NaCl, pH 7 in a range of 4 to 100 °C with a path length of 0**.5 mm on a** JASCO J-810 spectropolarimeter from 250 nm to 200 nm. Data fitted to a two state unfolding (m1+(m2-m1)/(1+(m0/m3)^m4)equation[4, 5, 6] which calculated the melting temperature of BfrWT – 49.7 ^{0}C, BfrCC-C-term – 50.1^{0}C and BfrCC-N/C-term – 45.5 ^{0}C. Data fitted using Kaleidagraph(Synergy).

BfrWT

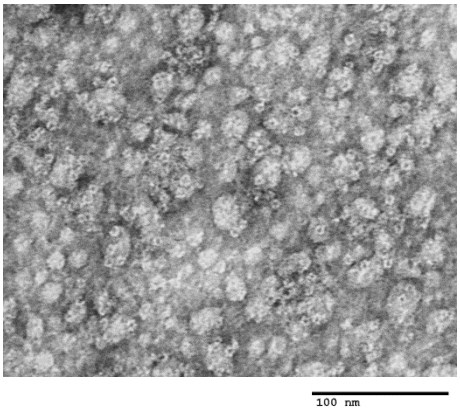


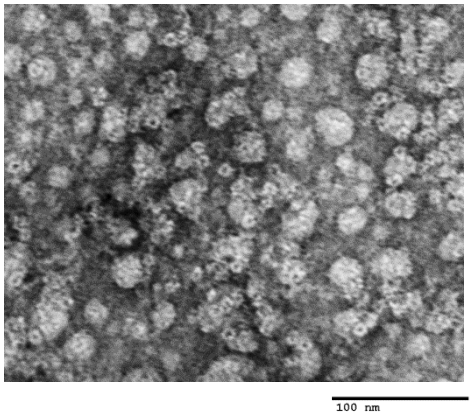
BfrCCPGCC-N-term



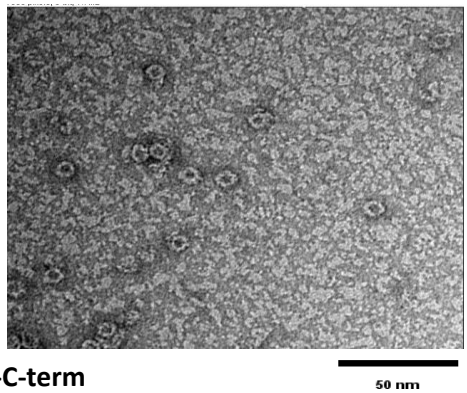


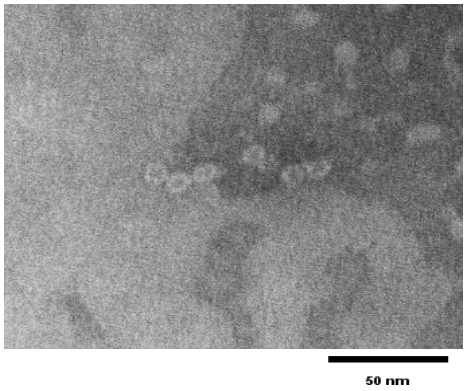
BfrCCPGCC-C-term



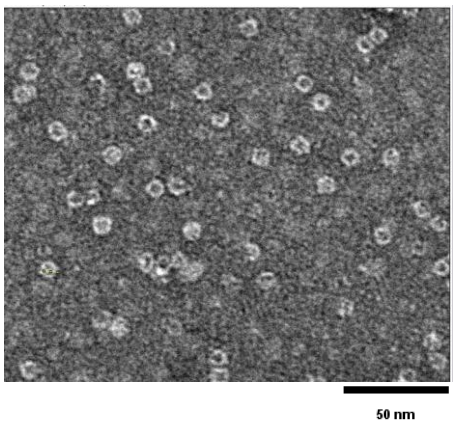


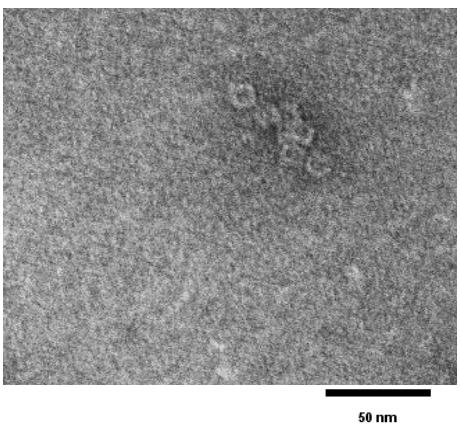
BfrCC-N/C-term



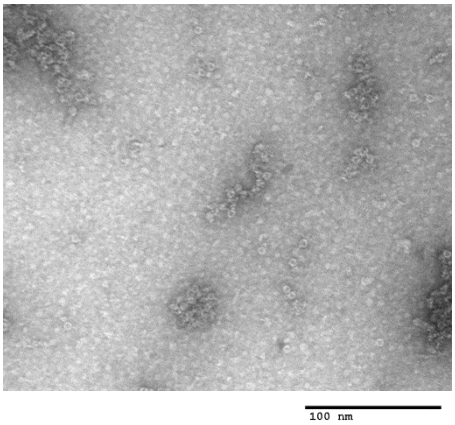


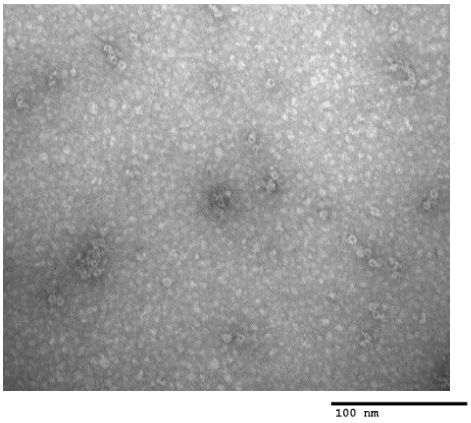
BfrCC-C-term



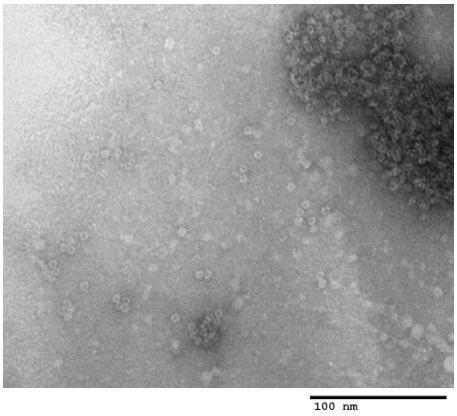


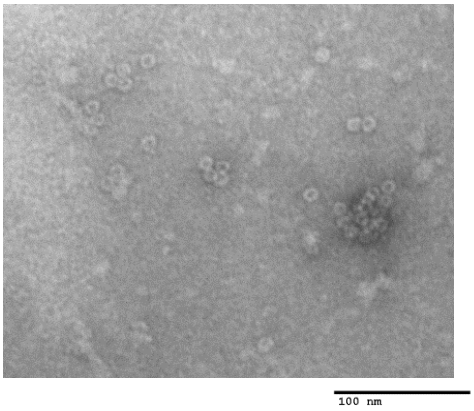
BfrCC-C-termN23F



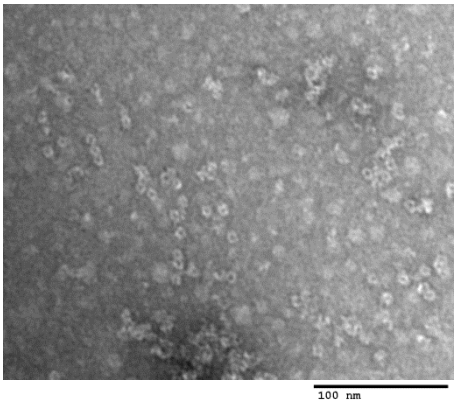


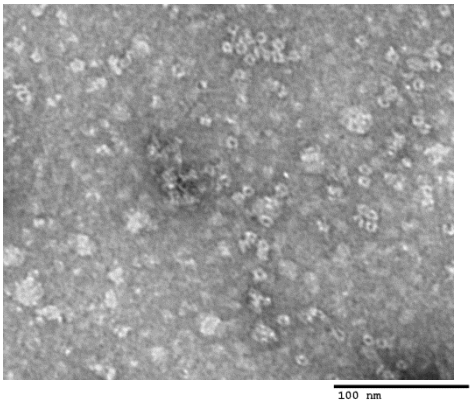
BfrCC-C-termD118F



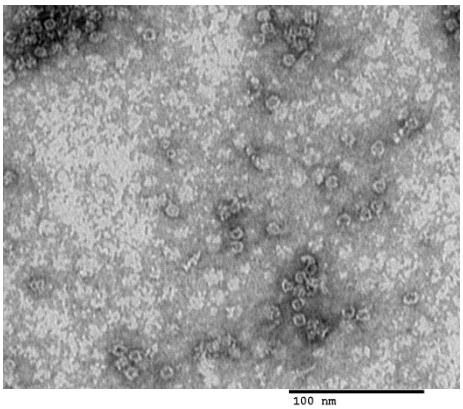


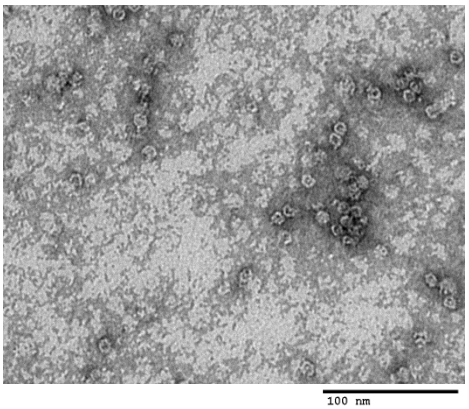
BfrCCPGCC-C-termN23F





BfrCCPGCC-C-termD118F

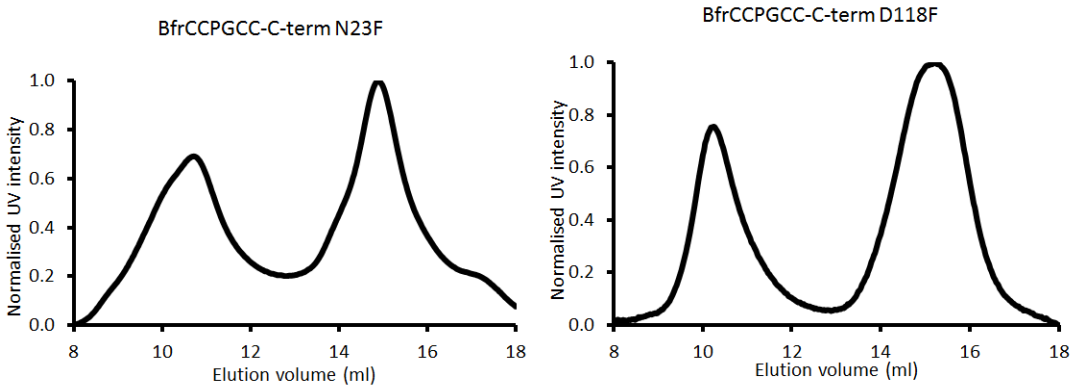




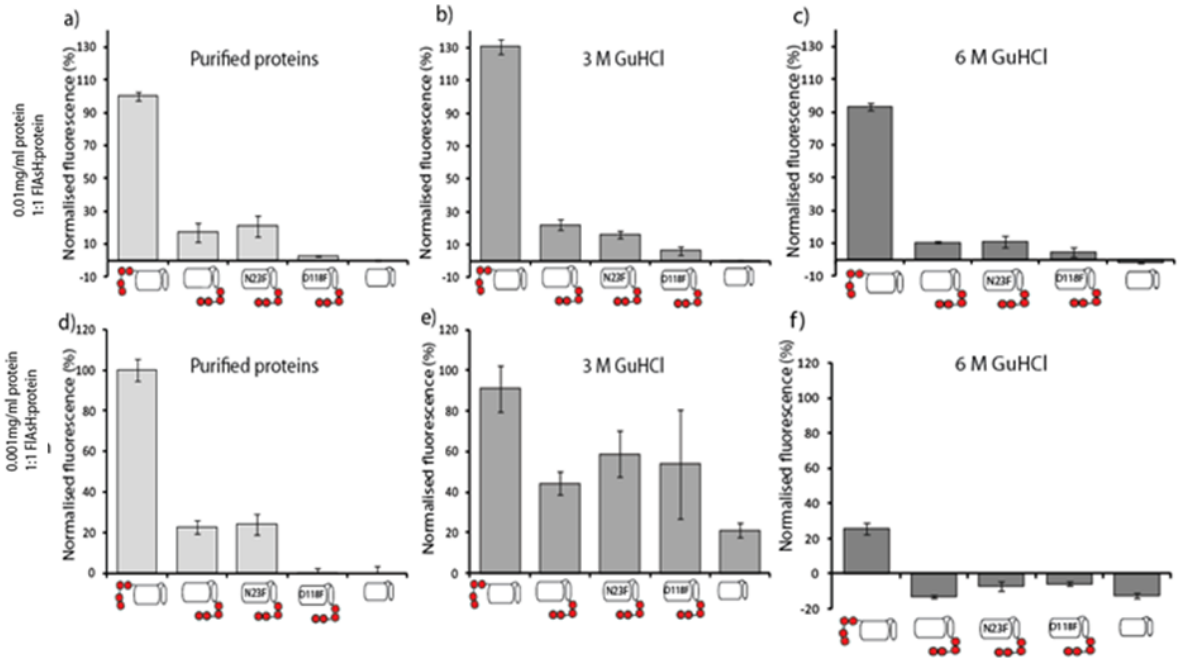
Appendix 4.8 – Transmission electron microscopy micrographs to show the self-assembly ability of all the controls and designs used in Chapter 4, using 1% uranyl acetate as a negative stain. Images made using ImageJ[7].

Protein	Measured mass (Da)	Expected Mass (Da)	Difference (Da)
BfrWT	18493.6	18495.0	1.4
BfrCCPGCC-N-term	19056.3	19061.8	5.5
BfrCCPGCC-C-term	19055.6	19061.8	6.2
BfrCC-C-term	18706.5	18701.3	5.2
BfrCC-N/C-term	18902.1	18907.6	5.5
BfrCC-C-termN23F	18741.2	18734.4	6.8
BfrCC-C-termD118F	18726.8	18733.4	6.6
BfrCCPGCC-C-term N23F	19058.0	19094.6	36.6
BfrCCPGCC-C-term D118F	19072.3	19093.6	21.3

Appendix 4.9 – Electrospray mass spectrometry analysis of all the Bfr controls and designs used in Chapter 4 that were purified, using 1% formic acid.



Appendix 4.10 – Size exclusion chromatogram of purified Bfr inaccessible binding site nanocage stabilised designs BfrCCPGCC-C-termN23f and BfrCCPGCC-C-termD118F. Chromatogram normalised to their highest intensities and are an average of three runs each (0.5 ml injection of 0.5 mg/ml protein (followed at 280 nm), using running buffer, 100 mM Tris.HCl, 100 mM NaCl, pH 7.8 at a flow rate of 0.5 ml/min.



Appendix 4.11 - Fluorescence analysis to detect Bfr assembly through binding site inaccessibility while varying the concentration of protein and denaturant but all with a 1:1 molar ratio of protein to fluorophore. a, b and c), 0.01 mg/ml protein concentration. d, e and f), 0.001 mg/ml protein concentration. a and d) Purified proteins only (light grey), b and e) purified proteins with 3 M Gu·HCl denaturant (grey) c and f) Purified proteins with 6 M Gu·HCl denaturant (Dark grey). All designs in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) with, 3.5 mM TCEP, 1 mM EDT, 1 mM 2-ME and varying concentrations of FlAsH-EDT_2 denaturant and protein.Each experiment repeated 6 times and error bars are S.D.

Appendix for Chapter 5

Table 1

Primer name	Primer sequence (5'-3')
Dps p22 F	GGA ACT CAT ATG AGC ACC GCA AAA CTG GTG
Dps p22 R	GGT ACT CTC GAT TTA TTC GAT GTT GCA TTC G

Table 2

Primer name	Primer sequence (5'-3')
DpsCC F	GAA TGC AAC ATC GAA TGC TGC TGA TTC TCC TCA ACC ATG GCG ATA TC
DpsCC R	GAT ATC GCC ATG GTT GAG GAG AAT CAG CAG CAT TCG ATG TTG CAT TC
DpsGCC F	GAA TGC AAC ATC GAA GGA TGC TGC TGA TCC TCA ACC ATG GCG ATA TC
DpsGCC R	GAT ATC GCC ATG GTT GAG GAT CAG CAG CAT CCT TCG ATG TTG CAT TC
DpsGGCC	GAA TGC AAC ATC GAA GGA GGT TGC TGC TGA TCA ACC ATG GCG ATA TC
DpsGGCC	GAT ATC GCC ATG GTT GAT CAG CAG CAA CCT CCT TCG ATG TTG CAT TC
DpsGGGCC	GAA TGC AAC ATC GAA GGC GGT GGA TGC TGC TGA ACC ATG GCG ATA TC
DpsGGGCC R	GAT ATC GCC ATG GTT CAG CAG CAT CCA CCT CCT TCG ATG TTG CAT TC
DpsPAGCC F	GAA TGC AAC ATC GAA CCA GCG GGA TGC TGC TGA ACC ATG GCG ATA TC
DpsPAGCC R	GAT ATC GCC ATG GTT CAG CAA CAT CCC GCT GGT TCG ATG TTG CAT TC
DpsCCPGCC F	GAA TGC AAC ATC GAA TGC TGC CCA GGT TGC TGC TAA ATG GCG ATA TC
DpsCCPGCC R	GAT ATC GCC ATT TAG CAG CAA CCT GGG CAG CAT TCG ATG TTG CAT TC

Table 3

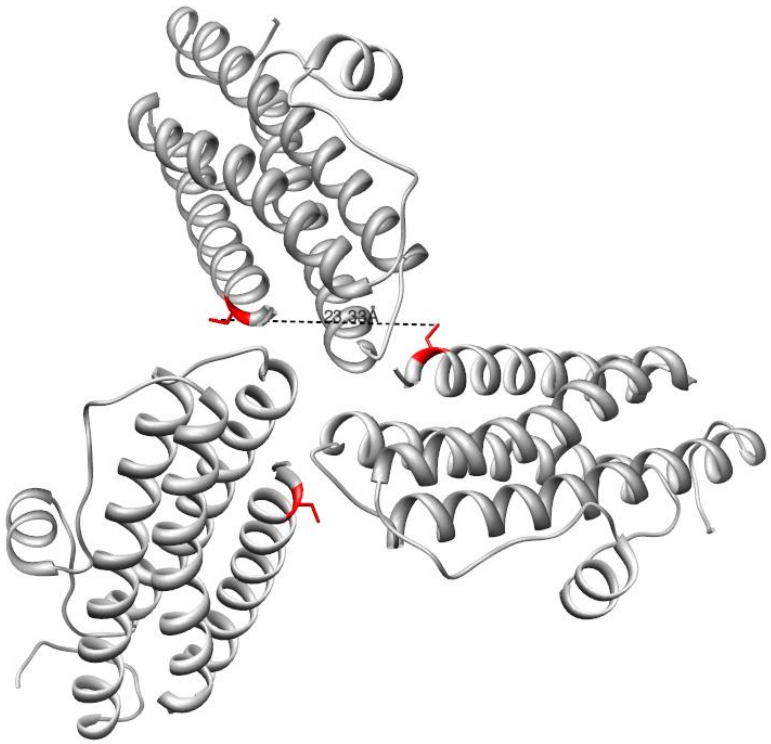
Primer name	Primer sequence (5'-3')
DpsR18C,N19C F	CAA CCA ACC TGC TGT ACA CCT GCT GCG ATG TGA GCG ATA GCG AAA AAA AAG
DpsR18C,N19C R	CTT TTT TTT CGC TAT CGC TCA CAT CGC AGC AGG TGT ACA GCA GGT TGG TTG
DpsN19C,D20C F	CCA ACC TGC TGT ACA CCC GTT GCT GCG TGA GCG ATA GCG AAA AAA AAG C
DpsN19C,D20C R	GCT TTT TTT TCG CTA TCG CTC ACG CAG CAA CGG GTG TAC AGC AGG TTG G
DpsR18C,D20C F	CAA CCA ACC TGC TGT ACA CCT GCA ACT GCG TGA GCG ATA GCG AAA AAA AAG
DpsR18C,D20C R	CTT TTT TTT CGC TAT CGC TCA CGC AGT TGC AGG TGT ACA GCA GGT TGG TTG
DpsD123C,A126C F	GAT CAT CTG AAA GAA CTG GCA TGC CGT TAC TGC ATC GTG GCA AAC GAT GTG CG
DpsD123C,A126C R	CGC ACA TCG TTT GCC ACG ATG CAG TAA CGG CAT GCC AGT TCT TTC AGA TGA TC
DpsA126C,N130C F	GAA CTG GCA GAT CGT TAC TGC ATC GTG GCA TGC GAT GTG CGT AAA GCA ATC GG
DpsA126C,N130C R	CCG ATT GCT TTA CGC ACA TCG CAT GCC ACG ATG CAG TAA CGA TCT GCC AGT TC

Table 4

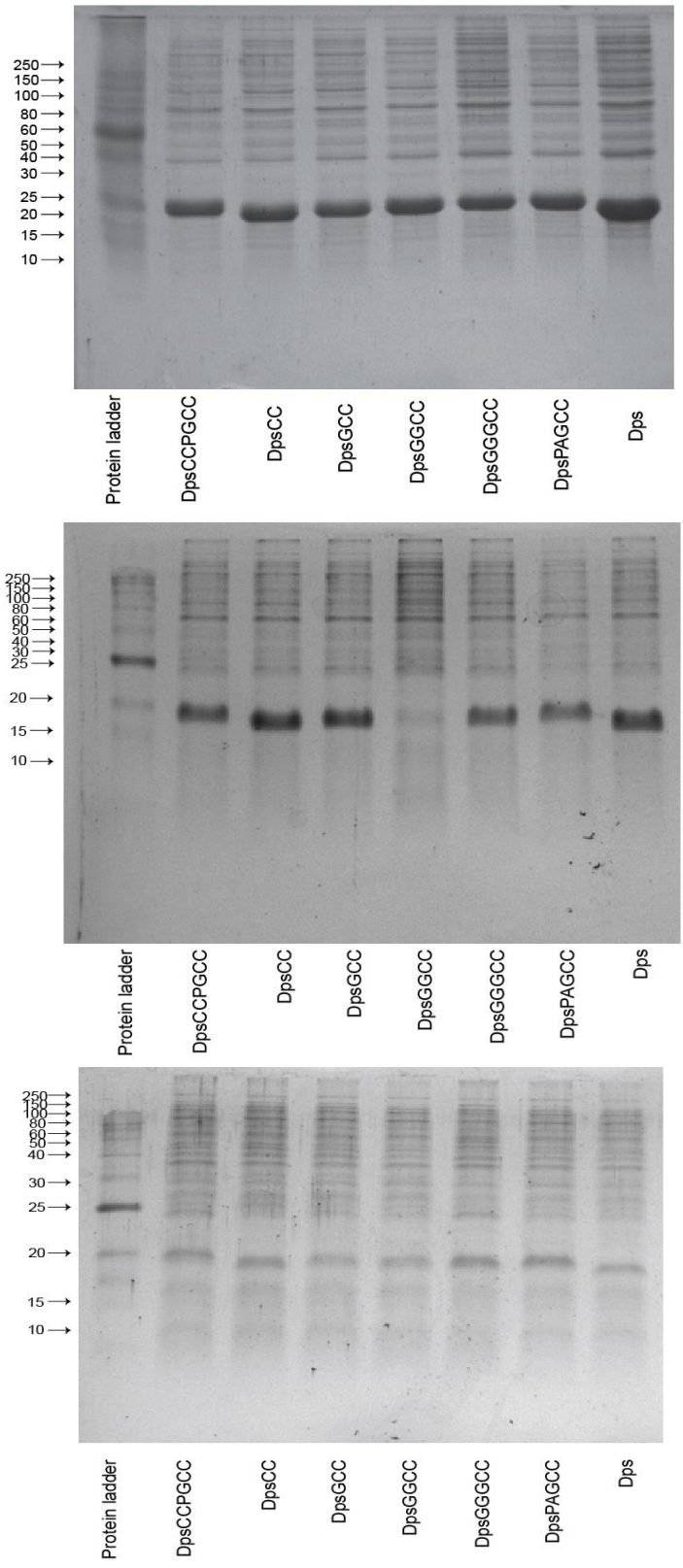
Primer name	Primer sequence (5'-3')
DpsGGGCC p32 R	GAG GAG AAG CCC GGT CAG CAG CAT CCA CCT CCT TC
DpsPAGCC p32 R	GAG GAG AAG CCC GGT CAG CAA CAT CCC GCT G
DpsCCPGCC p32 R	GAG GAG AAG CCC GGT TAG CAG CAA CCT GGG CAG

Dps F	GAC GAC GAC AAG ATG AGC ACC GCA AAA CTG GTG

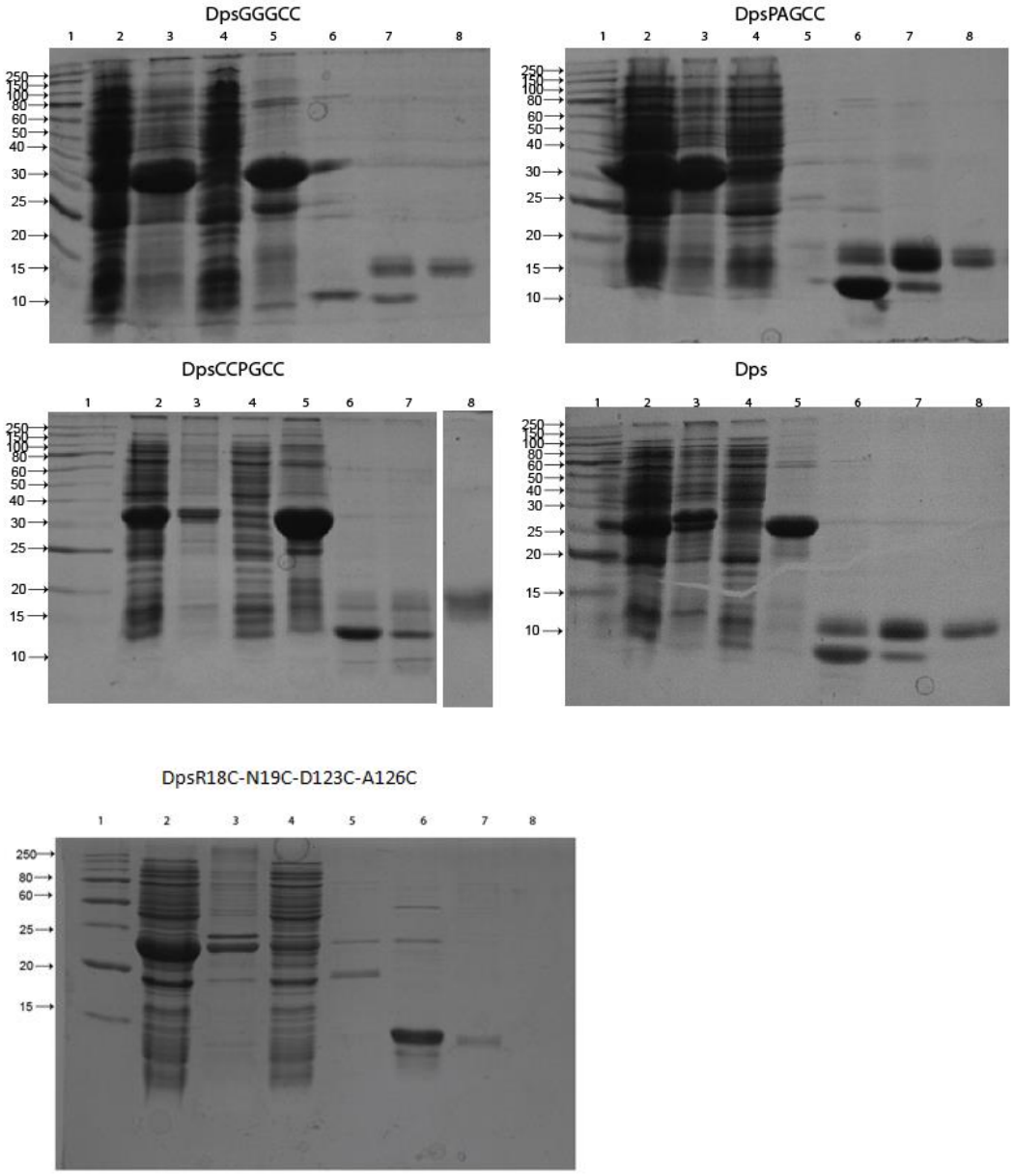
Appendix 5.1 - Primers used for the construction of all designs and controls of Dps used in Chapter 5. Table 1 – Primers used to transfer the wild type Dps gene from pET-32b to pET-22b. Table 2 – Primers used to create all bipartite Dps designs. Table 3 – Primers used to create all geometrically correct Dps designs. Table 4 – Primers used to transfer the successful Dps bipartite designs from pET-22b to pET-32b for purification. Primers sourced from Supranom, Singapore. F – Forward primer, R – reverse primer.



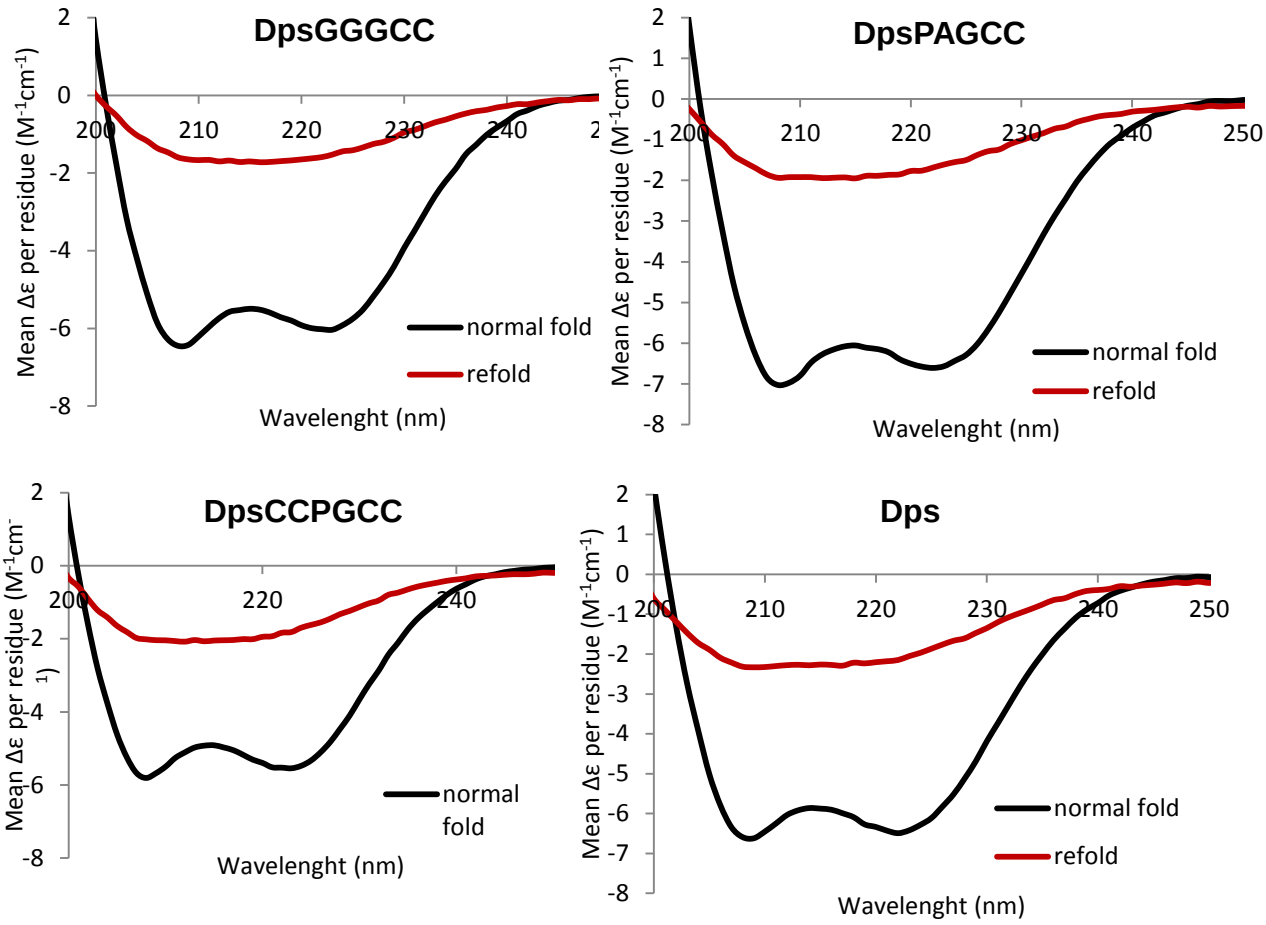
Appendix 5.2 – Three fold axis of symmetry of Dps (PDB:1DPS[1]) highlighting native cysteine residues in the structure (red). While there is only one per-structure and near the C-terminus, their distance a is 23.33 Å apart (from sulphur to sulphur atom). Image made using Chimera[8].



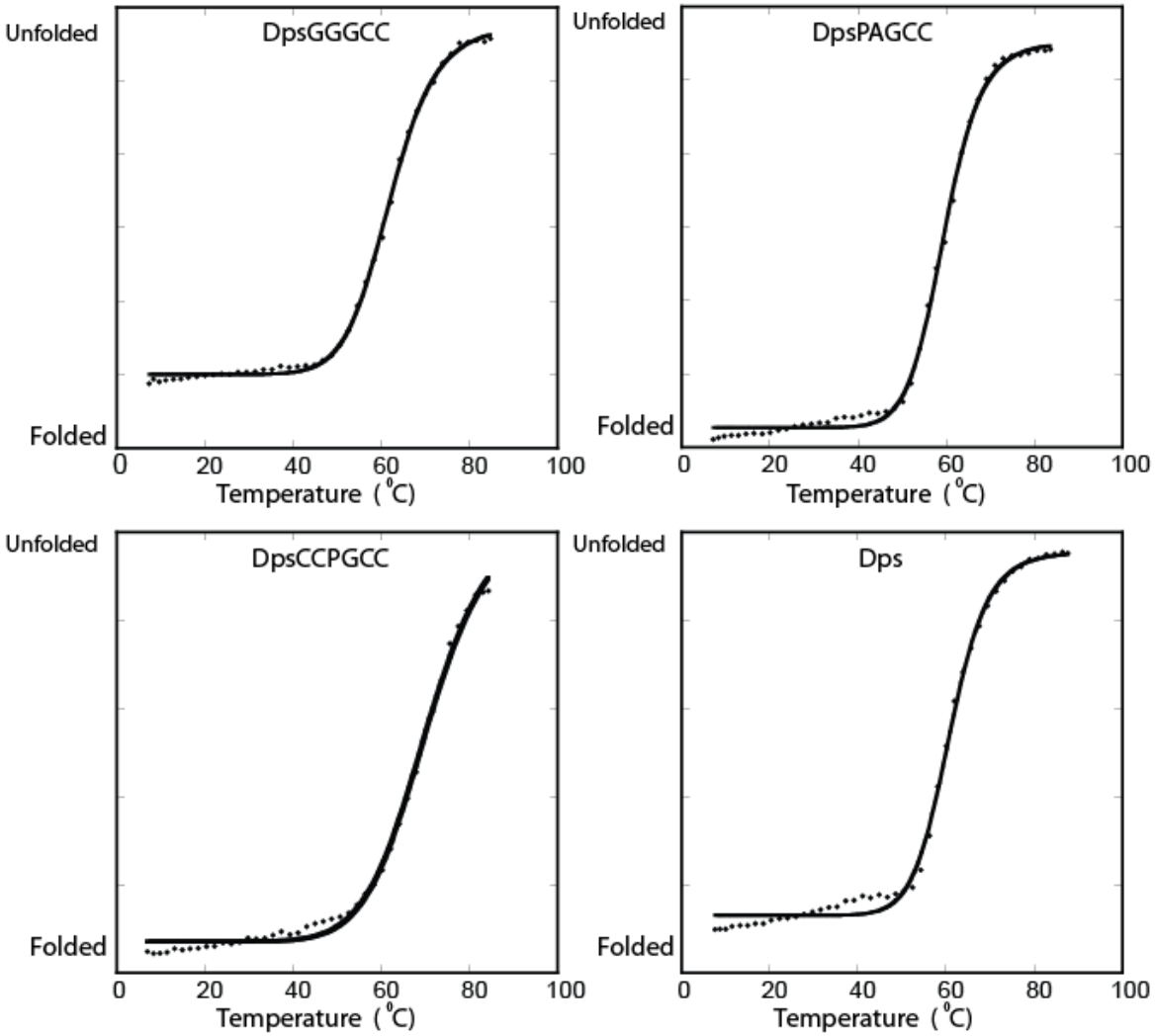
Appendix 5.3 – SDS-PAGE (15%) of lysate samples for each Dps design diluted to 1 mg/ml total to evaluate the expression level of each design. Each design was expressed three times.



Appendix 5.4 – SDS-PAGE gel (15%) for each of the purified proteins explored in Chapter 5. Lane 1: Protein standard ladder (NEB), Lane 2: Soluble fraction, lane 3: insoluble fraction, lane 4: wash from first histrap, lane 5: elution from first histrap, lane 6: elution from second histrap after EK digestion, lane 7: wash from second histrap after EK digestion, lane 8: SEC.



Appendix 5.5 – Circular Dichroism spectra demonstrating the effect of heating on the fold of the Dps controls and bipartite designs purified in Chapter 5. Black: CD signal before heating denaturation. Red: CD signal after cooling post heating denaturation. The purified proteins (0.2 mg/ml) were inspected in FlAsH buffer (100 mM Tris, 100 mM NaCl, 1 mM EDTA, pH 7.8). This experiment was performed on an Applied Photophysics LTD Chirascan spectrometer in a range of 200 nm to 260 nm with a path length of 0.5 mm.



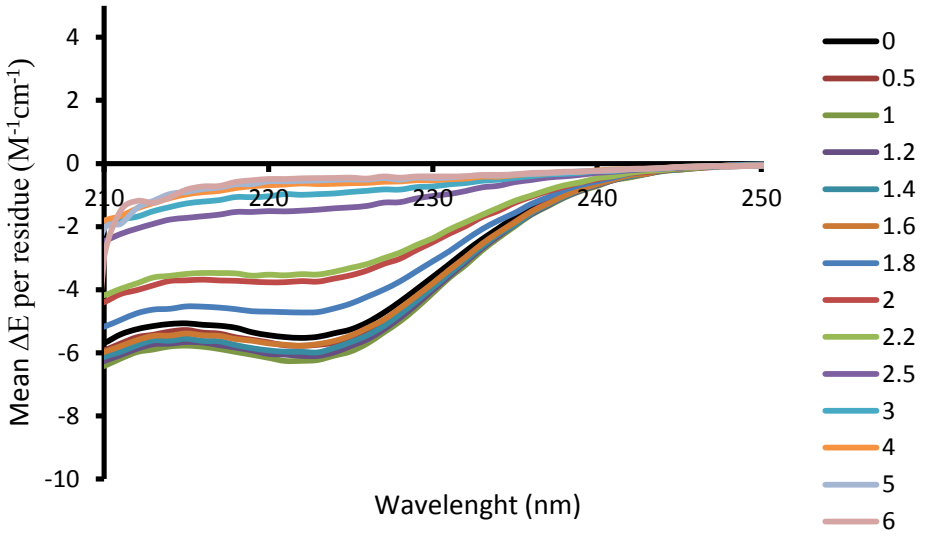
Appendix 5.6 – Thermal melting plots of the Dps controls and bipartite designs purified in Chapter 5 at 222 nm in CD. Thermal melts were performed on all purified proteins (0.2 mg/ml in FlAsH buffer, 100 mM Tris, 100 mM NaCl, 1 mM EDTA, pH 7.8) in a range of 4 to 85 °C with a path length of 0**.5 mm** on an Applied Photophysics LTD Chirascan spectrometer from 250 nm to 200 nm. Data fitted to a two state unfolding (m1+(m2-m1)/(1+(m0/m3)^m4)equation[4, 5, 6] which calculated the melting temperature (see below). Data fitted using Kaleidagraph(Synergy).

Design	Tm average (°C)
DpsCCPGCC	70.35 ± 1.42
DpsGGGCC	61.87 ± 0.13
DpsPAGCC	58.84 ± 0.40
Dps	61.66 ± 1.49

Appendix 5.7 – Melting temperature averages from the data seen in Appendix 5.6. S.D from three heating denaturation's from each protein.

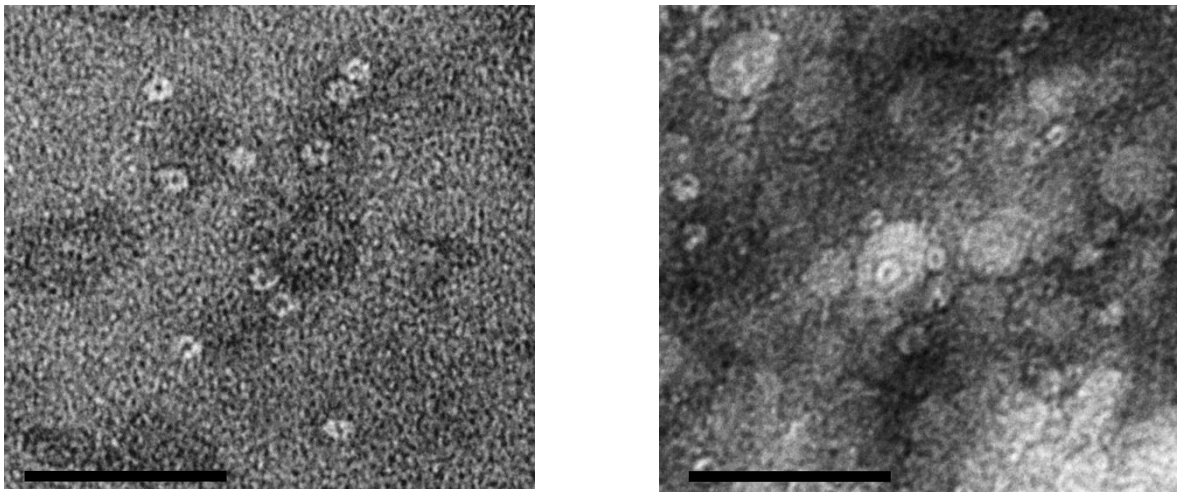
Protein	Measured mass (Da)	Expected mass (Da)	Difference (Da)
DpsCCPGCC	19273.0	19278.0	5.0
DpsGGGCC	19100.3	19088.8	-11.5
DpsPAGCC	19154.0	19142.8	-11.2
Dps	18710.2	18711.3	1.1

Appendix 5.8 – Electrospray mass spectrometry analysis of all the Dps controls and designs that were purified in Chapter 5, using 1% formic acid.

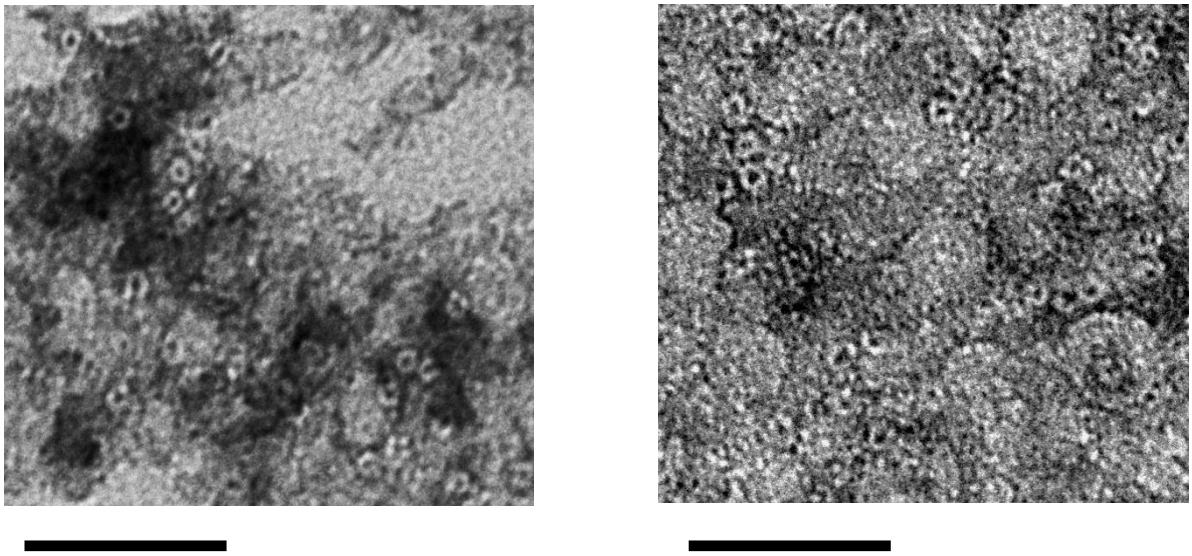


Appendix 5.9 – CD signal from the Dps bipartite design DpsPAGCC during denaturation with Gu·HCl. The purified proteins (0.2 mg/ml) were inspected in FlAsH buffer (100 mM Tris, 100 mM NaCl, 1 mM EDTA, pH 7.8). This experiment was performed on an Applied Photophysics LTD Chirascan spectrometer in a range of 200 nm to 260 nm with a path length of 0.5 mm. The concentration of Gu·HCl was varied and is recorded in the data (M).

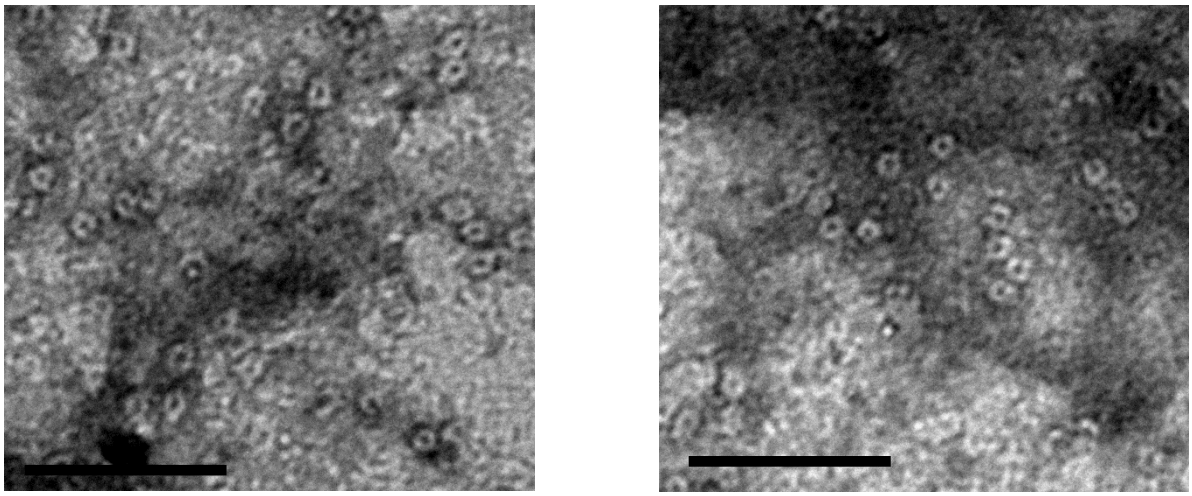
DpsGGGCC



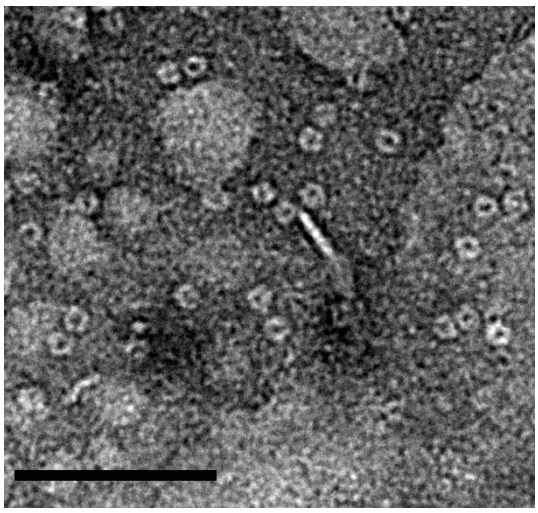
DpsPAGCC

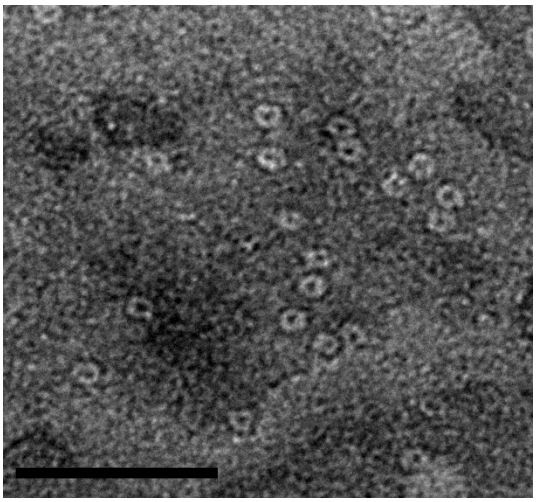


DpsCCPGCC



Dps





Appendix 5.10 – Transmission electron microscopy micrographs to show the self-assembly ability of all the controls and designs used in Chapter 5, using 1% uranyl acetate as a negative stain. Images made using ImageJ[7].

Appendix for Chapter 6

Table 1

Name	5’-3’ sequence
Dps LIC F	GAC GAC GAC AAG ATG AGC ACC GCA AAA CTG GTG
DpsCCPGCC LIC R	GAG GAG AAG CCC GGT TAG CAG CAA CCT GGG CAG
Dps PAGCC LIC R	GAG GAG AAG CCC GGT CAG CAA CAT CCC GCT G
DpsWT LIC R	GAG GAG AAG CCC GGT TAT TCG ATG TTG CAT TCG

Table 2

Name	5’-3’ sequence
DpsPAGCC (R83A) F	CTG GAT ACC ATG GCA GAA GCA GCA GTG CAG CTG GGT GGT
DpsPAGCC (R83A) R	ACC ACC CAG CTG CAC TGC TGC TTC TGC CAT GGT ATC CAG
DpsPAGCC (R133A) F	TCG TGG CAA ACG ATG TGG CAA AAG CAA TCG GTG AAG C
DpsPAGCC (R133A) R	GCT TCA CCG ATT GCT TTT GCC ACA TCG TTT GCC ACG A
DpsPAGCC (R83D) F	CTG GAT ACC ATG GCA GAA GAC GCA GTG CAG CTG GGT GGT
DpsPAGCC (R83D) R	ACC ACC CAG CTG CAC TGC GTC TTC TGC CAT GGT ATC CAG
DpsPAGCC (R133D) F	TCG TGG CAA ACG ATG TGG ACA AAG CAA TCG GTG AAG C
DpsPAGCC (R133D) R	GCT TCA CCG ATT GCT TTG TCC ACA TCG TTT GCC ACG A
DpsPAGCC (R83W) F	CTG GAT ACC ATG GCA GAA TGG GCA GTG CAG CTG GGT GGT
DpsPAGCC (R83W) R	ACC ACC CAG CTG CAC TGC CCA TTC TGC CAT GGT ATC CAG
DpsPAGCC (R133W) F	TCG TGG CAA ACG ATG TGT GGA AAG CAA TCG GTG AAG C
DpsPAGCC (R133W) R	GCT TCA CCG ATT GCT TTC CAC ACA TCG TTT GCC ACG A

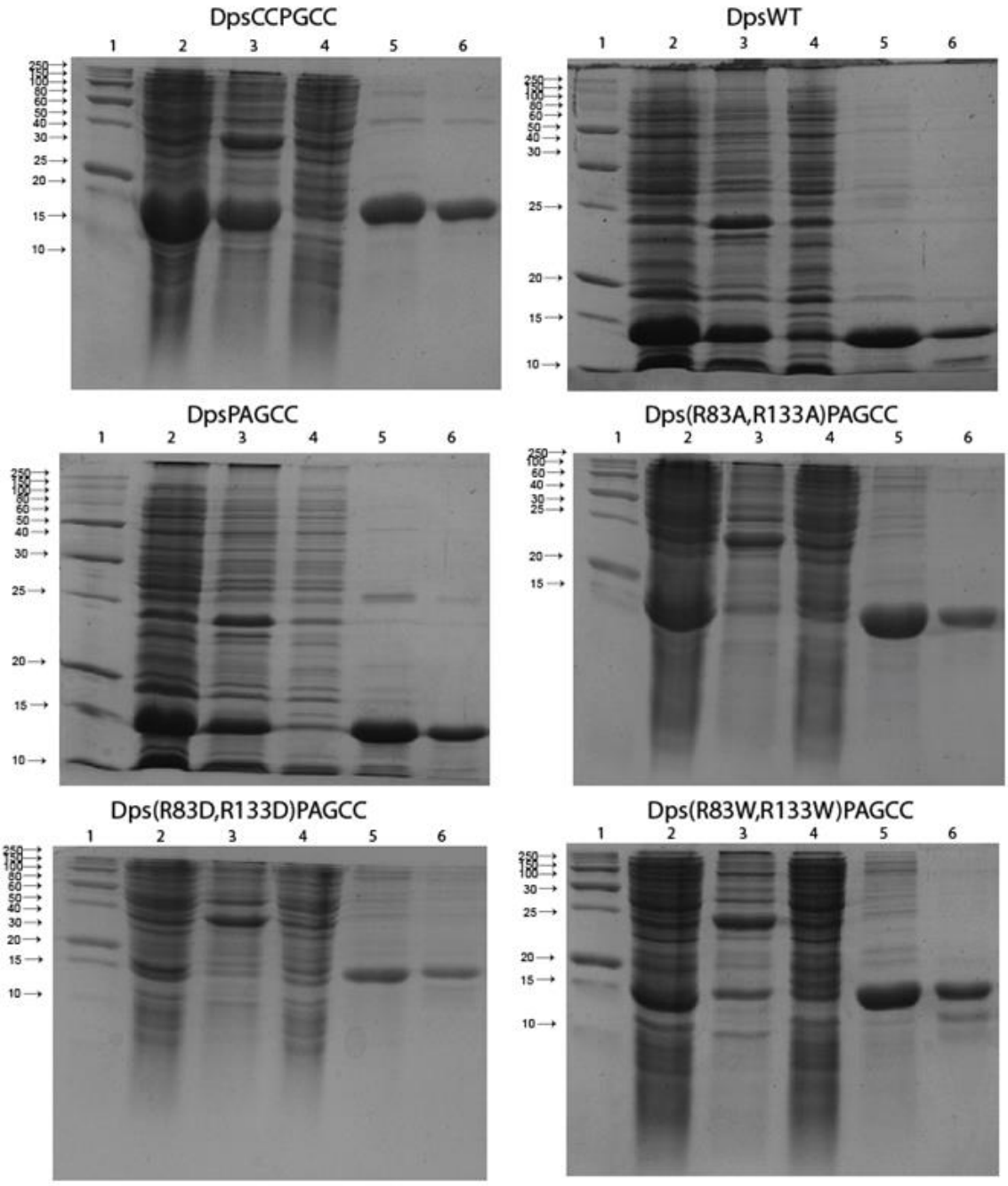
Table 3

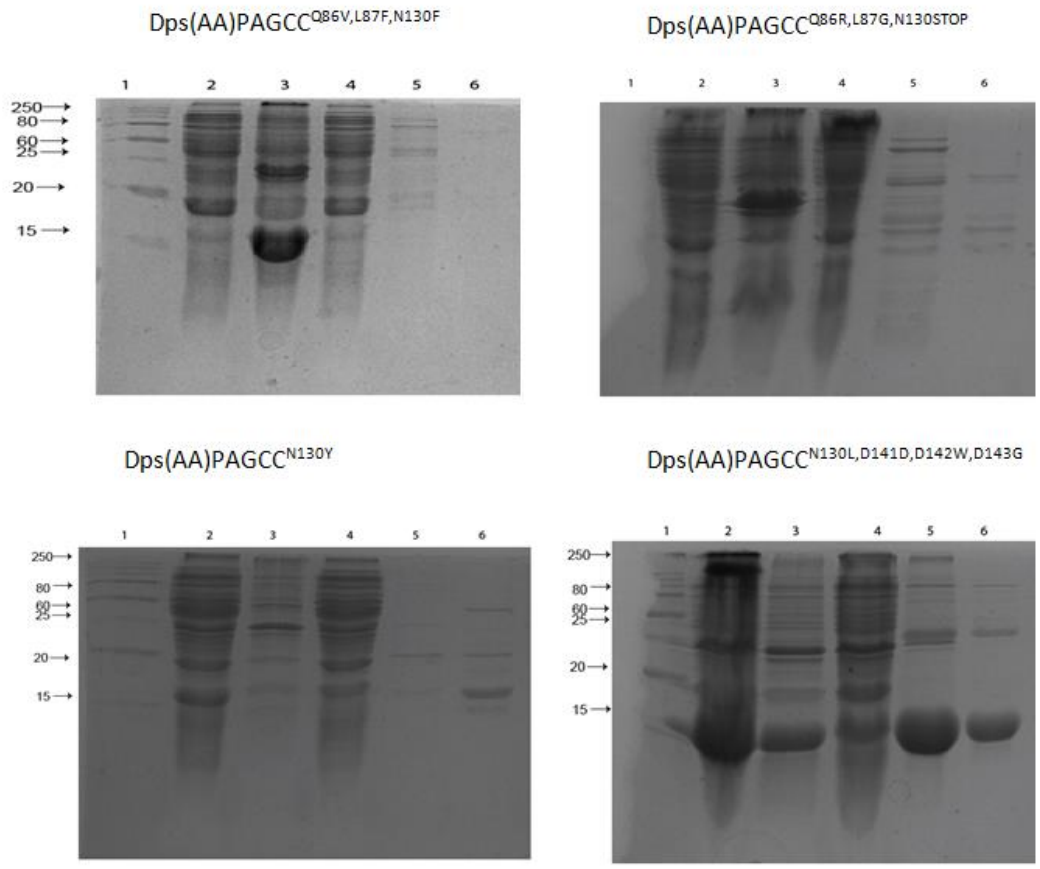
Name	5’-3’ sequence	TM (°C)
Dps(AA)PAGCC 19, 20, 21, F	TAC ACC CGT NNS NNS NNS AGC GAT AG	64
Dps(AA)PAGCC 86, 87, F	CAG CAG TGN NSN NSG GTG GTG T	64
Dps(AA)PAGCC 130, F	ATC GTG GCA NNS GAT GTG GCA AAA G	63
Dps(AA)PAGCC 141, 142, 143, F	GTG AAG CAA AAN NSN NSN NSA CCG CAG AT	64
Dps(AA)PAGCC 152, 153, 156, F	CAG CAN NSN NSG ATC TGN NSA AAT TCC T	63
DpsPAGCC R	GCA ACA TCC CGC TGG TTC GAT GTT	63

Table 4

Primer name	Primer sequence (5'-3')	TM (°C)
Dps(AA)PAGCC 19, 20, 21, F	CCT GCT GTA CAC CCG TNN SNN SNN SAG CGA TAG CGA AAA AAA AG	66
Dps(AA)PAGCC 86, 87, F	AGA AGC AGC AGT GNN SNN SGG TGG TGT GGC ACT	71
Dps(AA)PAGCC 130, F	ATC GTT ACG CAA TCG TGG CAN NSG ATG TGG CAA AAG CAA TC	67
Dps(AA)PAGCC 141, 142, 143, F	ATC GGT GAA GCA AAA NNS NNS NNS ACC GCA GAT ATC CTG ACC G	66
Dps(AA)PAGCC 152, 153, 156, F	GAT ATC CTG ACC GCA GCA NNS NNS GAT CTG NNS AAA TTC CTG TG	67
DpsPAGCC R	TCA GCA ACA TCC CGC TGG TTC GAT GTT GCA TTC GAT GAA	68

Appendix 6.1 – Primers used in the investigation undertaken in Chapter 6. Table 1 – Primers used to transfer the Dps genes from Chapter 5 that were in pET-32b to pET-46. Table 2 – Primers used to introduce the nanocage crippling mutations into the Dps positions R83 and R133. Table 3 – Short primers used in the creation of a TPCR library. Table 4 – Long primers used in the creation the TPCR library. Primer sourced from Eurofins, Germany. F – Forward primer, R – reverse primer.


DpsCCPGCC
1 2 3 4 5 6
250
150
100
80
60
50
40
30
25
20
15
10
DpsWT
1 2 3 4 5 6
250
150
100
80
60
50
40
30
25
20
15
10
DpsPAGCC
1 2 3 4 5 6
250
150
100
80
60
50
40
30
25
20
15
10
Dps(R83A,R133A)PAGCC
1 2 3 4 5 6
250
100
60
40
30
25
20
15
Dps(R83D,R133D)PAGCC
1 2 3 4 5 6
250
150
100
80
60
50
40
30
20
15
10
Dps(R83W,R133W)PAGCC
1 2 3 4 5 6
250
150
100
30
25
20
15
10

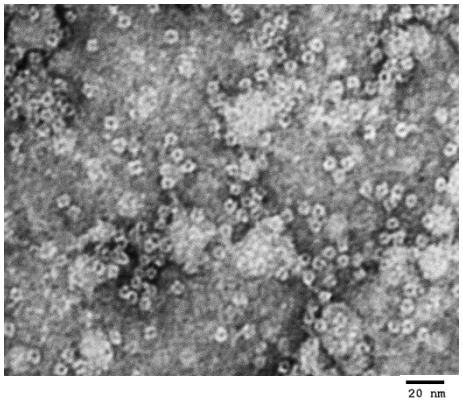


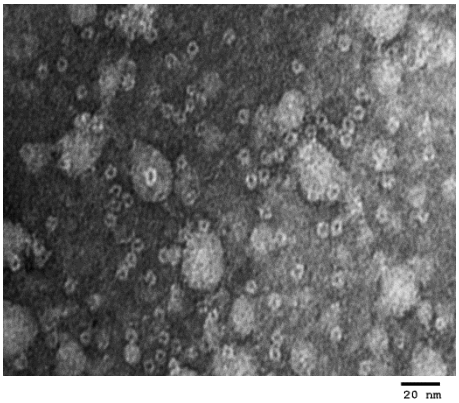
Appendix 6.2 – SDS-PAGE gel (15%) for each of the purified proteins explored in Chapter 6. Lane 1: Protein standard ladder (NEB), Lane 2: Soluble fraction, lane 3: insoluble fraction, lane 4: wash from histrap, lane 5: elution from histrap, lane 6: SEC.

Protein	Measured mass (Da)	Expected Mass (Da)	Difference (Da)
DpsWT	20710.9	20422.3	-288.5
DpsPAGCC	20710.8	20853.9	143.1
Dps(AA)PAGCC	20540.7	20681.6	140.9
DpsCCPGCC	20833.4	21214.3	380.9
Dps(AA)PAGCC $^{Q86V,L87F, N130F}$	20824.5	20719.7	-104.8
Dps(AA)PAGCC $^{Q86R, L87G, N130Stop}$	16994.7	16208.8	-785.9
Dps(AA)PAGCC ^{N130Y}	20828.4	20730.7	-97.7
Dps(AA)PAGCC $^{N130L, D1414D, D142W, D143G}$	20540.7	20695.8	155.1
Dps(DD)PAGCC	20650.0	20769.6	119.6
Dps(WW)PAGCC	20798.0	20911.8	113.8

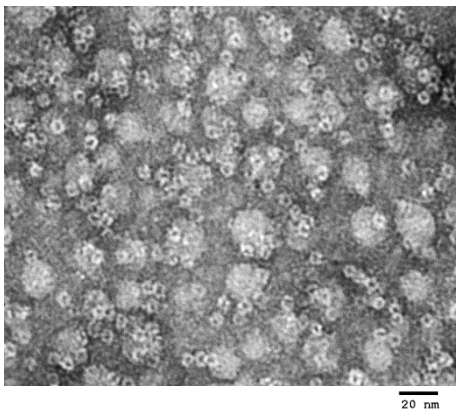
Appendix 6.3 – Maldi-TOF mass spectrometry analysis of the purified proteins explored in Chapter 6.

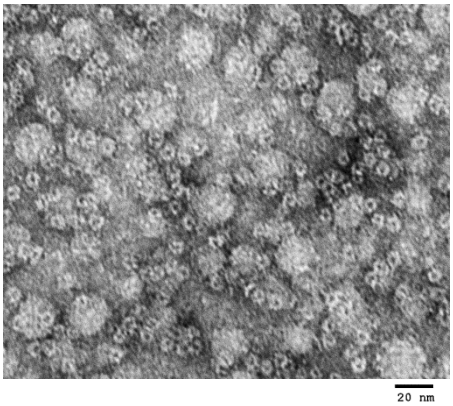
DpsWT



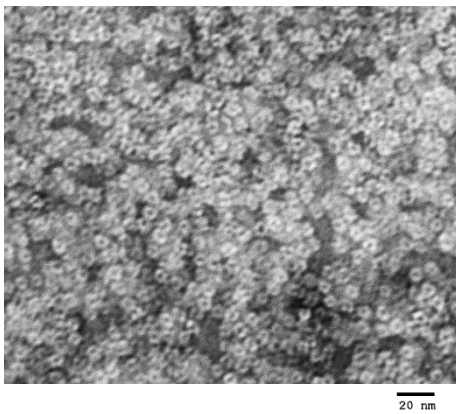


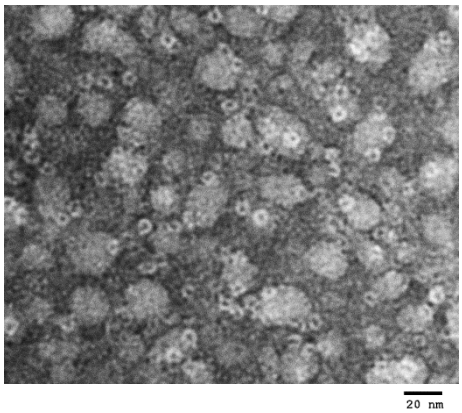
DpsPAGCC



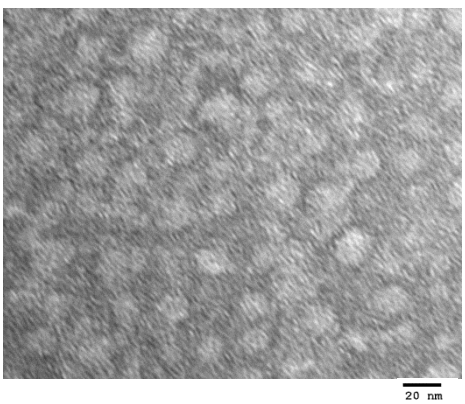


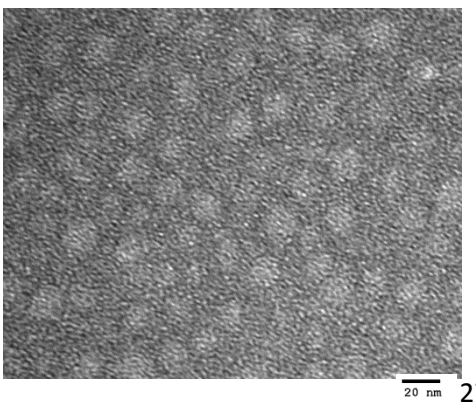
DpsCCPGCC



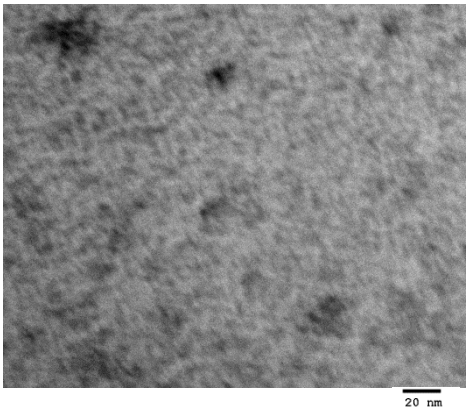


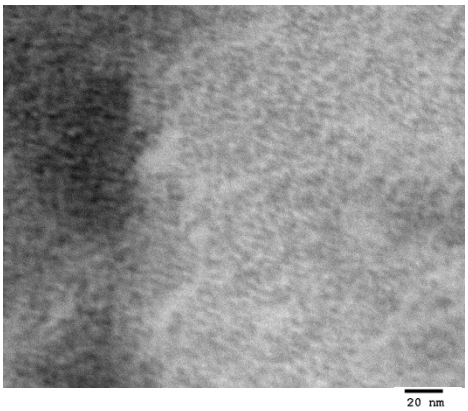
Dps(AA)PAGCC



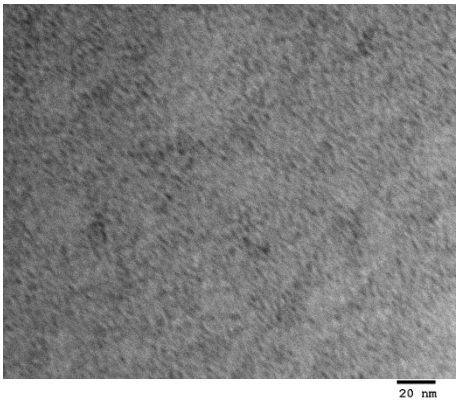


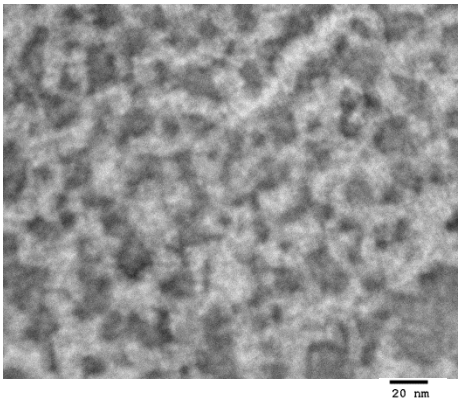
Dps(DD)PAGCC



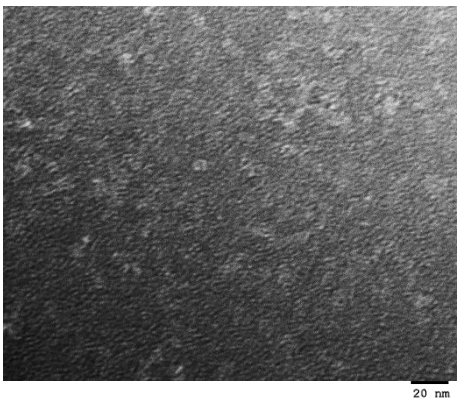


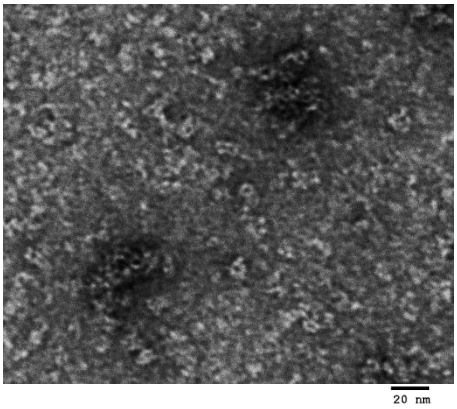
Dps(WW)PAGCC



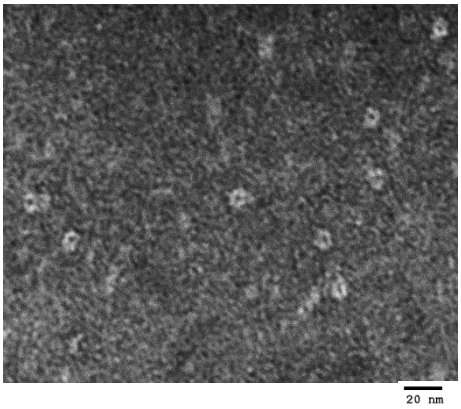


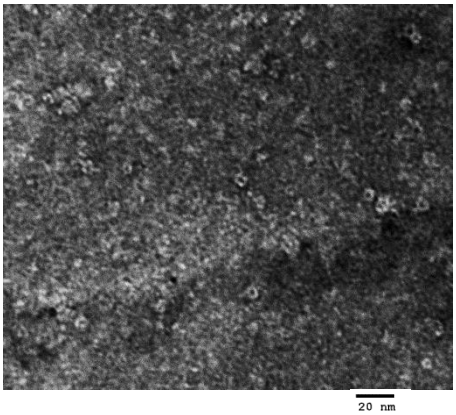
Dps(AA)PAGCC [Q86V, L87F, N130F]



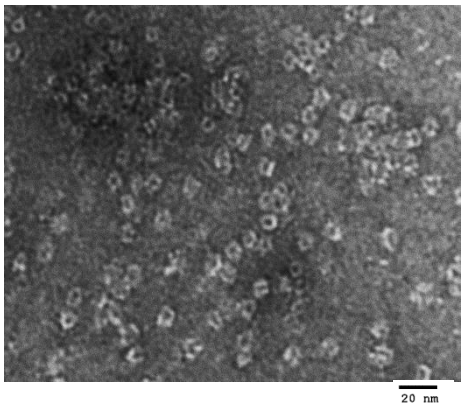


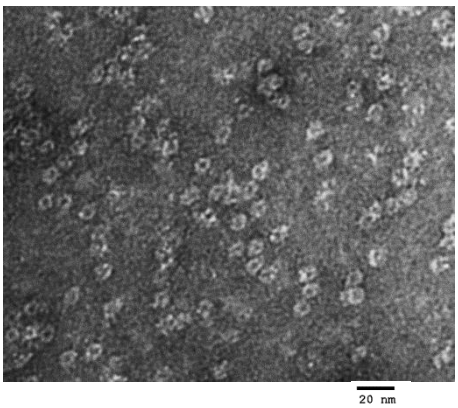
Dps(AA)PAGCC ^{N130Y}



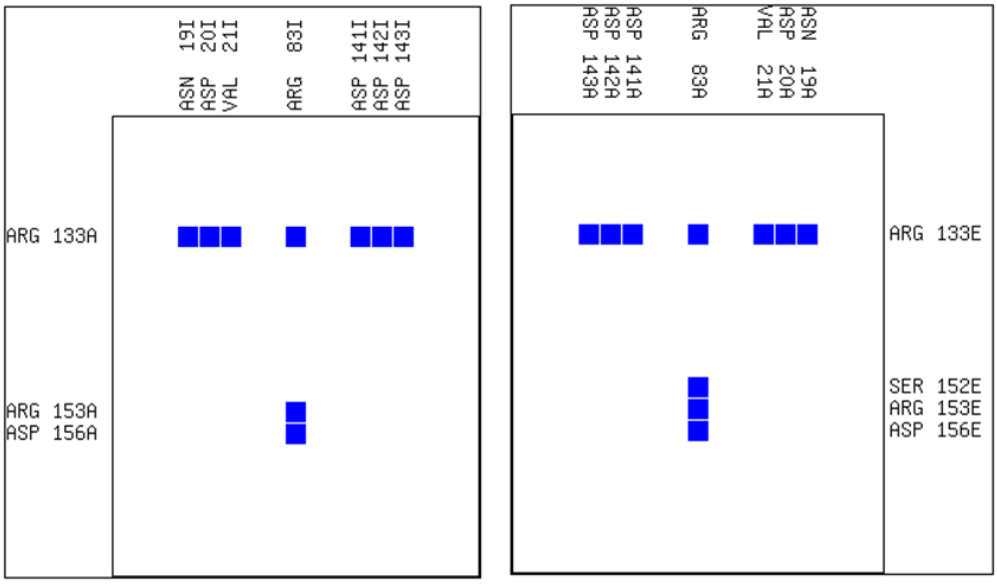


Dps(AA)PAGCC $^{N130L, D141D, D142W, D143F}$

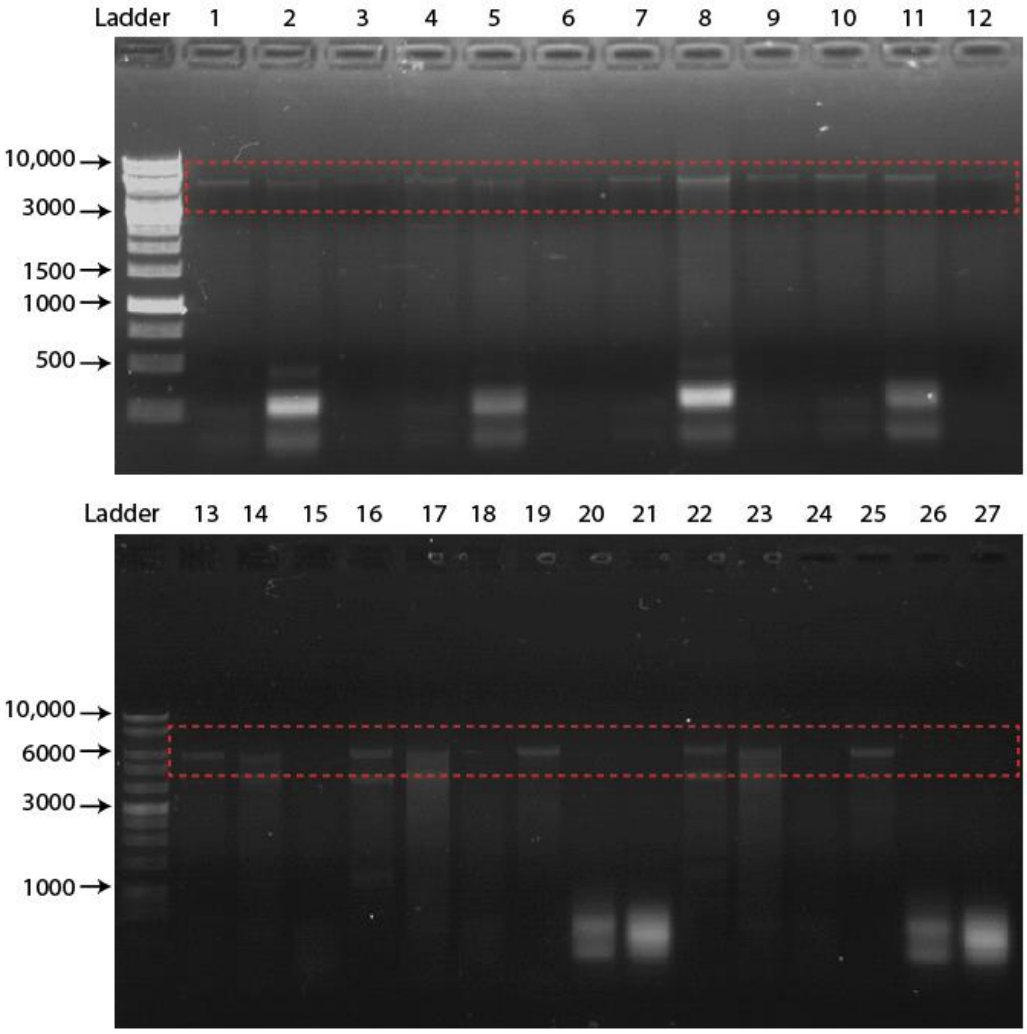


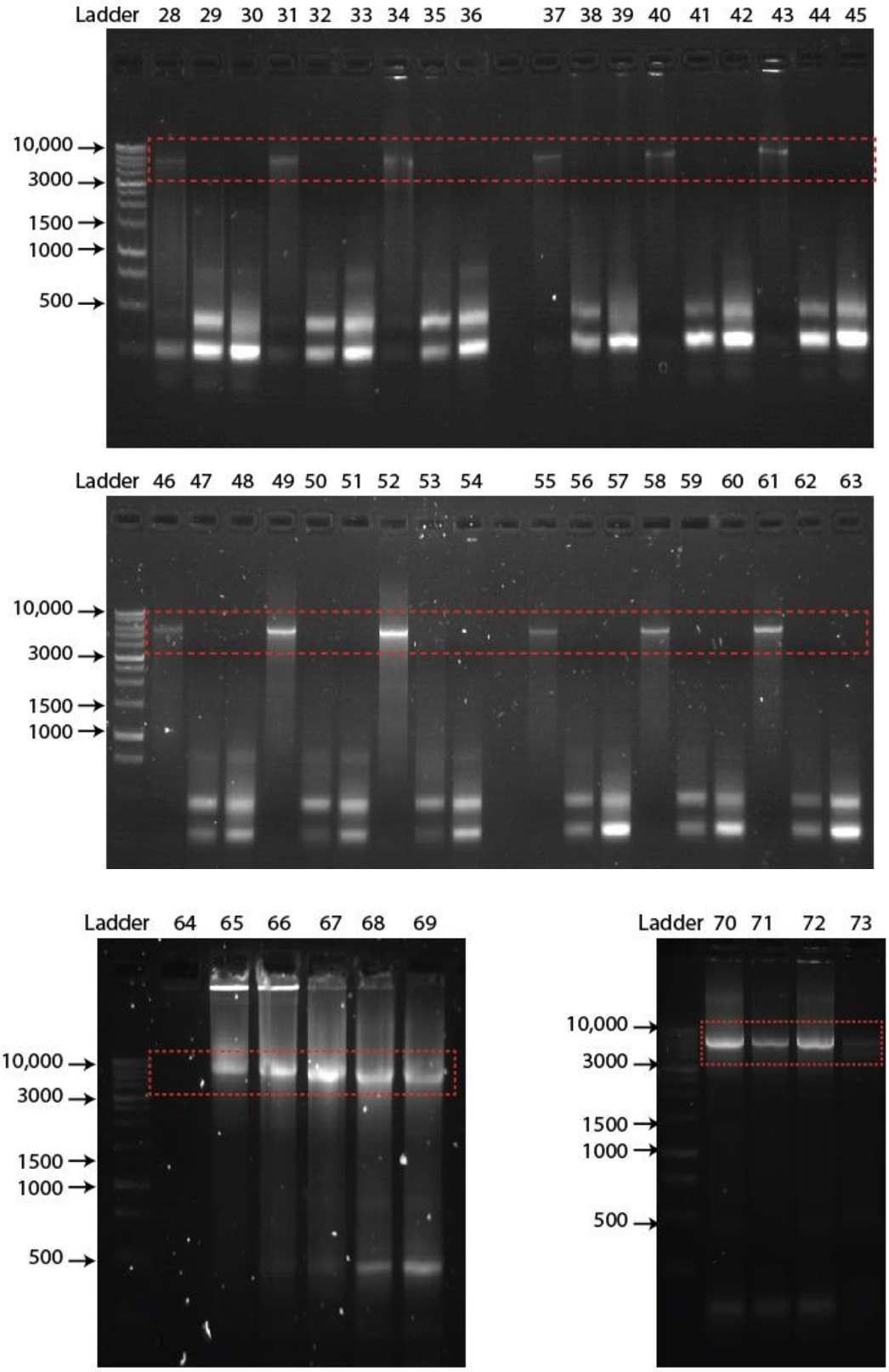


Appendix 6.4 – Transmission electron microscopy micrographs to show the self-assembly ability of all the controls and designs used in Chapter 6, using 1% uranyl acetate as a negative stain. Images made using ImageJ[7].



Appendix 6.5 – Analysis of the interaction network at the three-fold Dps like symmetry axis of a crystal structure (PDB:1DPS[1]) around the residues R82 and R133. Blue box represents an interaction between those amino acids. Analysis performed using the Contact map analysis (CMA) tool of the SPACE software[9].


Ladder 1 2 3 4 5 6 7 8 9 10 11 12
10,000
3000
1500
1000
500
Ladder 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27
10,000
6000
3000
1000



Appendix 6.6 – 1% agarose gels to analyse the success of the PCR reaction to create the TPCR library, see Appendix 6.7 for PCR conditions of each gel lane. The first lane in each gel is of a 10

kb ladder. All wells loaded with 10 µl of PCR product solution after mixing with 6 x loading dye. Red dashed box highlights the position of the expected amplified plasmid band.

Sample	Primers used	Amount of primer (ng) Primer F1	Primer F2	Primer F3	Primer F4	Primer F5	Primer R1	Plasmid	DMSO (%)	Program
1	Long	40	40	40	40	40	40	20	2	1
2	Long	80	80	80	80	80	80	20	2	1
3	Long	20	20	20	20	20	20	20	2	1
4	Long	40	40	40	40	40	40	20	2	2
5	Long	80	80	80	80	80	80	20	2	2
6	Long	20	20	20	20	20	20	20	2	2
7	Long	40	40	40	40	40	40	20	2	3
8	Long	80	80	80	80	80	80	20	2	3
9	Long	20	20	20	20	20	20	20	2	3
10	Long	40	40	40	40	40	40	20	2	4
11	Long	80	80	80	80	80	80	20	2	4
12	Long	20	20	20	20	20	20	20	2	4
13	Short	10	10	10	10	10	10	20	2	1
14	Short	20	20	20	20	20	20	20	2	1
15	Short	40	40	40	40	40	40	20	2	1
16	Short	10	10	10	10	10	10	20	2	3
17	Short	20	20	20	20	20	20	20	2	3
18	Short	40	40	40	40	40	40	20	2	3
19	Long	10	10	10	10	10	10	20	2	3
20	Long	20	20	20	20	20	20	20	2	3
21	Long	40	40	40	40	40	40	20	2	3
22	Short	10	10	10	10	10	10	20	2	4
23	Short	20	20	20	20	20	20	20	2	4
24	Short	40	40	40	40	40	40	20	2	4
25	Long	10	10	10	10	10	10	20	2	4
26	Long	20	20	20	20	20	20	20	2	4
27	Long	40	40	40	40	40	40	20	2	4
28	Long	5	5	5	5	5	5	20	0	3
29	Long	10	10	10	10	10	10	20	0	3
30	Long	15	15	15	15	15	15	20	0	3
31	Long	5	5	5	5	5	5	20	1	3
32	Long	10	10	10	10	10	10	20	1	3
33	Long	15	15	15	15	15	15	20	1	3
34	Long	5	5	5	5	5	5	20	2	3
35	Long	10	10	10	10	10	10	20	2	3
36	Long	15	15	15	15	15	15	20	2	3

37	Long	5	5	5	5	5	5	20	0	4
38	Long	10	10	10	10	10	10	20	0	4
39	Long	15	15	15	15	15	15	20	0	4
40	Long	5	5	5	5	5	5	20	1	4
41	Long	10	10	10	10	10	10	20	1	4
42	Long	15	15	15	15	15	15	20	1	4
43	Long	5	5	5	5	5	5	20	2	4
44	Long	10	10	10	10	10	10	20	2	4
45	Long	15	15	15	15	15	15	20	2	4
46	Long	5	5	5	5	5	5	10	0	3
47	Long	10	10	10	10	10	10	10	0	3
48	Long	15	15	15	15	15	15	10	0	3
49	Long	5	5	5	5	5	5	10	1	3
50	Long	10	10	10	10	10	10	10	1	3
51	Long	15	15	15	15	15	15	10	1	3
52	Long	5	5	5	5	5	5	10	2	3
53	Long	10	10	10	10	10	10	10	2	3
54	Long	15	15	15	15	15	15	10	2	3
55	Long	5	5	5	5	5	5	10	0	4
56	Long	10	10	10	10	10	10	10	0	4
57	Long	15	15	15	15	15	15	10	0	4
58	Long	5	5	5	5	5	5	10	1	4
59	Long	10	10	10	10	10	10	10	1	4
60	Long	15	15	15	15	15	15	10	1	4
61	Long	5	5	5	5	5	5	10	2	4
62	Long	10	10	10	10	10	10	10	2	4
63	Long	15	15	15	15	15	15	10	2	4
64	Long	0	0	0	0	0	0	20	2	4
65	Long	2	2	2	2	2	2	20	2	4
66	Long	4	4	4	4	4	4	20	2	4
67	Long	6	6	6	6	6	6	20	2	4
68	Long	8	8	8	8	8	8	20	2	4
69	Long	10	10	10	10	10	10	20	2	4
70	Long	10	6	6	6	10	6	20	2	4
71	Mix	Short 6	Long 6	Long 6	Long 6	Short 6	Long 6	20	2	4
72	Long	6	6	6	6	6	6	20	2	4
73	Short	6	6	6	6	6	6	20	2	4

Program 1 - The PCR reaction contained a total volume of 50 µl including, Pfu reaction buffer (Promega, 5 µl of 10), dNTP mix (Promega, 2 µl of a solution containing dATP, dTTP, dGTP, and dCTP at 2 mM each), and primers, template and DMSO concentrations as stated above and,

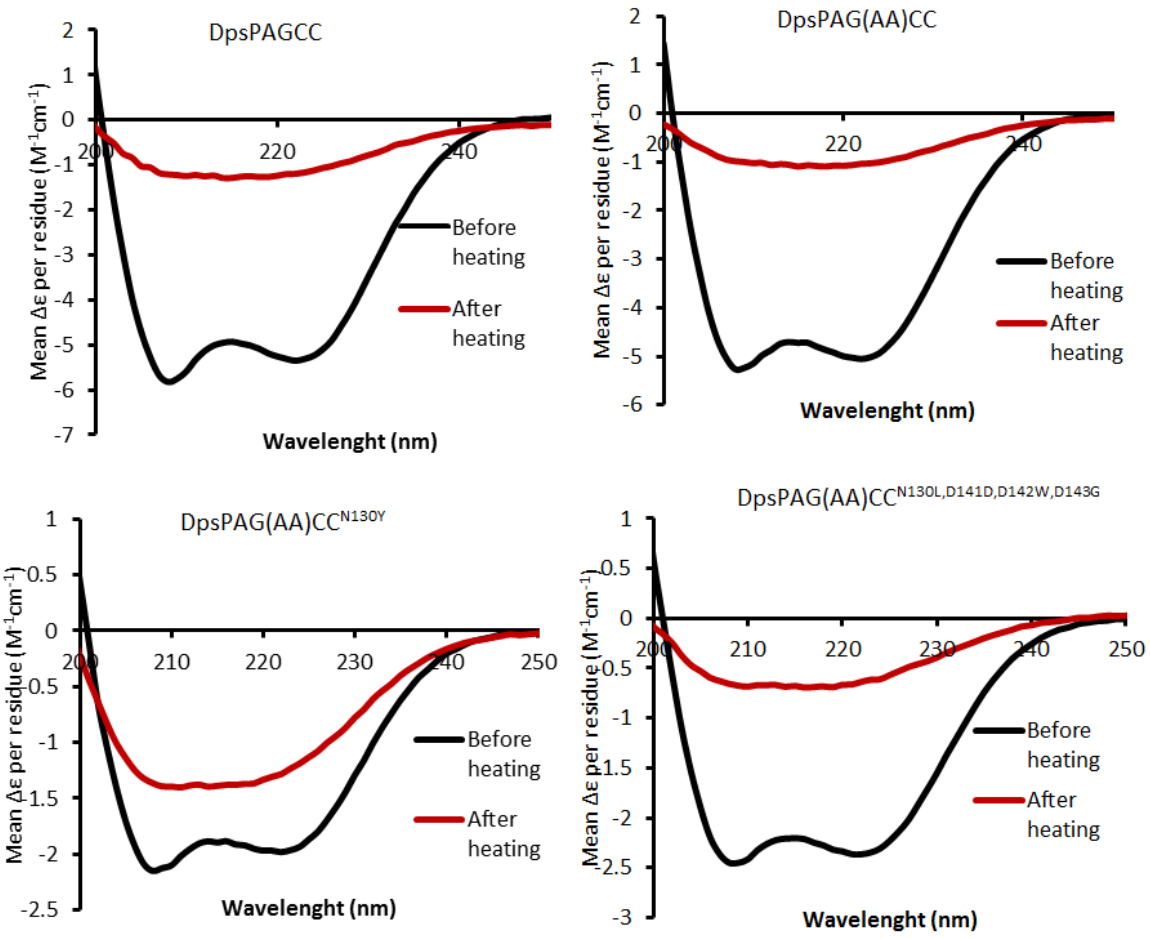
Pfu polymerase (Promega, 1 µl of 3 U/µl) and deionised H_2O (to a total of 50 µl) which was subjected to an initial melting step (95 °C for 5 min) followed by 20 cycles of amplification (95 °C for 30 s, 56 °C for 1 min, raising 0.3 °C every cycle and 72 °C for 6 min) followed by 20 more cycles of amplification (95 °C for 30 s, 64 °C for 1 min, raising 0.3 °C every cycle and 72 °C for 6 min) finishing with 72 °C for 10 min.

Program 2 - The PCR reaction contained a total volume of 50 µl including, Pfu reaction buffer (Promega, 5 µl of 10), dNTP mix (Promega, 2 µl of a solution containing dATP, dTTP, dGTP, and dCTP at 2 mM each), and primers, template and DMSO concentrations as stated above and, Pfu polymerase (Promega, 1 µl of 3 U/µl) and deionised H_2O (to a total of 50 µl) which was subjected to an initial melting step (95 °C for 5 min) followed by 10 cycles of amplification (95 °C for 30 s, 56 °C for 1 min and 72 °C for 6 min) with 10 more cycles of amplification (95 °C for 30 s, 58 °C for 1 min and 72 °C for 6 min) with 10 more cycles of amplification (95 °C for 30 s, 64 °C for 1 min and 72 °C for 6 min) with 10 more cycles of amplification (95 °C for 30 s, 66 °C for 1 min and 72 °C for 6 min) followed with 72 °C for 10 min.

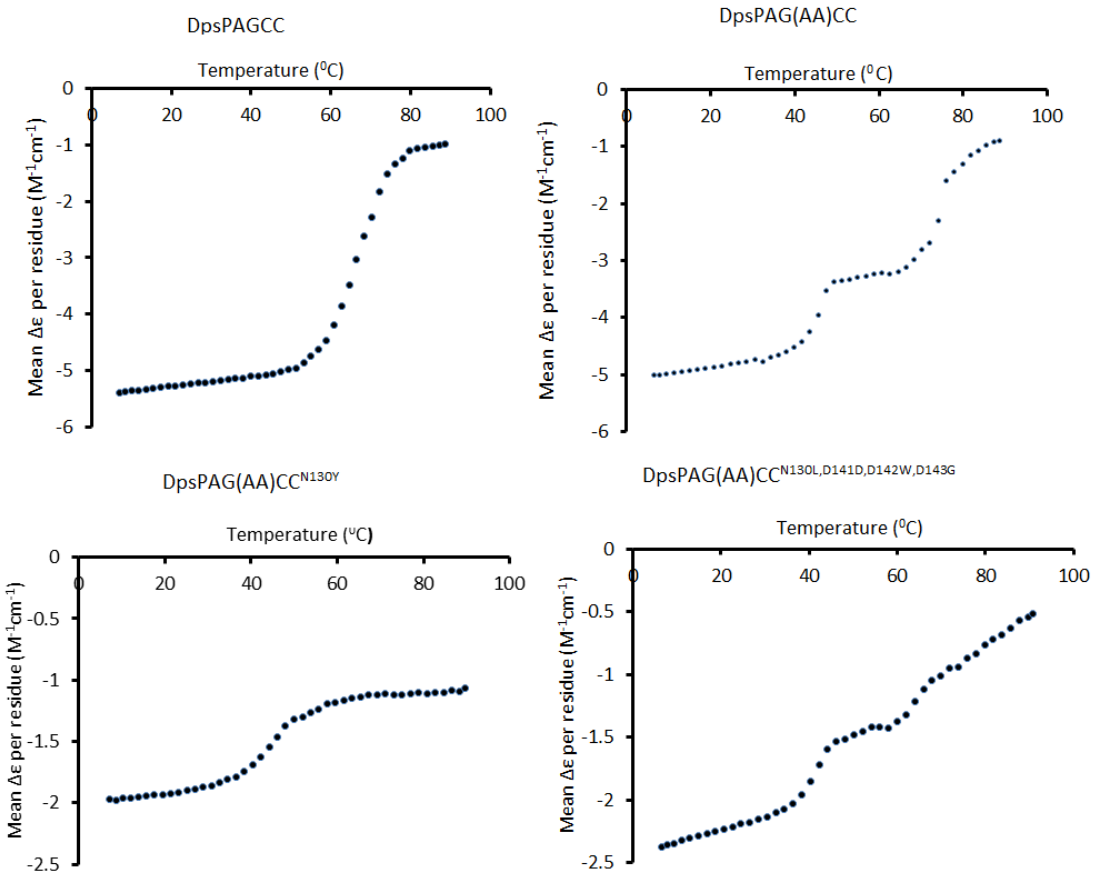
Program 3 - The PCR reaction contained a total volume of 50 µl including, Pfu reaction buffer (Promega, 5 µl of 10), dNTP mix (Promega, 2 µl of a solution containing dATP, dTTP, dGTP, and dCTP at 2 mM each), and primers, template and DMSO concentrations as stated above and, Pfu polymerase (Promega, 1 µl of 3 U/µl) and deionised H_2O (to a total of 50 µl) which was subjected to an initial melting step (95 °C for 5 min) followed by 10 cycles of amplification (95 °C for 30 s, 58 °C for 1 min and 72 °C for 6 min) with 10 more cycles of amplification (95 °C for 30 s, 60 °C for 1 min and 72 °C for 6 min) with 10 more cycles of amplification (95 °C for 30 s, 66 °C for 1 min and 72 °C for 6 min) with 10 more cycles of amplification (95 °C for 30 s, 68 °C for 1 min and 72 °C for 6 min) followed with 72 °C for 10 min.

Program 4 (see methods and materials Chapter 6)- The PCR reaction contained a total volume of 50 µl including, Pfu reaction buffer (Promega, 5 µl of 10), dNTP mix (Promega, 2 µl of a solution containing dATP, dTTP, dGTP, and dCTP at 2 mM each), and primers, template and DMSO concentrations as stated above and, Pfu polymerase (Promega, 1 µl of 3 U/µl) and deionised H_2O (to a total of 50 µl) which was subjected to an initial melting step (95 °C for 5 min) followed by 40 cycles of amplification (95 °C for 30 s, 58 °C for 1 min, raising 0.3 °C every cycle and 72 °C for 6 min) followed by 72 °C for 10 min.

Appendix 6.7 – Summary of the conditions assayed to create the TPCR library for Chapter 6. Table highlights the sample name, which corresponds to the gel lane in Appendix 6.6, the primer set that was used either long or short (see Appendix 6.1) the concentration of each primer, plasmid and DMSO as well as the thermal cycle program that was used.



Appendix 6.8 – Circular Dichroism spectra demonstrating the effect of heating on the fold of the Dps controls and bipartite designs purified in Chapter 6. Black: CD signal before heating denaturation. Red: CD signal after cooling post heating denaturation. The purified proteins (0.2 mg/ml) were inspected in FlAsH buffer (100 mM Tris, 100 mM NaCl, 1 mM EDTA, pH 7.8). This experiment was performed on an Applied Photophysics LTD Chirascan spectrometer in a range of 200 nm to 250 nm with a path length of 0.5 mm.



Appendix 6.9 – Thermal melting plots of the Dps controls and designs purified in Chapter 6 at 222 nm in CD. Thermal melts were performed on all purified proteins (0.2 mg/ml in FlAsH buffer, 100 mM Tris, 100 mM NaCl, 1 mM EDTA, pH 7.8) in a range of 4 to 85 °C with a path length of 0.5 mm on an Applied Photophysics LTD Chirascan spectrometer from 250 nm to 200 nm.

Sequence Appendix

Comparison of the sequencing results with expected sequence. Sequencing result shown on the top line, aligned with the expected sequence on the bottom. Yellow indicates extension or codon mutation. Protein design and vector used is indicated in each sequence alignment title.

Sequence Appendix 2 - Sequence alignment for all mutants used in nanoparticle formation in chapter 2. 'Query', obtained from sequence results from Supranom, Singapore, aligned against expected sequences 'Subject'. Aligned using Blast[10].

Wild type proteins in pET-32b

DpsWT in pET-32b

```
Query  9    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  68
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  60

Query  69   GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  128
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  129  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  188
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  189  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  248
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  249  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  308
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  309  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  368
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  369  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  428
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  420

Query  429  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  488
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  489  TTCATCGAATGCAACATCGAATAA  512
            ||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAATAA  504
```

BfrWT in pET-32b

```
Query  543   ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  602
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1     ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  60

Query  603   GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  662
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Sbjct  61    GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  120

Query  663   AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  722
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121   AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  180

Query  723   CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  782
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181   CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  240

Query  783   GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  842
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241   GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  300

Query  843   CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  902
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301   CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  360

Query  903   ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  962
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361   ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  420

Query  963   ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGTTGA  1019
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421   ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGTTGA  477
```

Sequence Appendix 3 - Sequence alignment for the Dps+E mutant used in chapter 3. E-helix highlighted in grey. 'Query', obtained from sequence results from Supranom, Singapore, aligned against expected sequences 'Subject'. Aligned using Blast[10].

Dps+E in pET-46 plasmid

```
Query  61   ATGGCACATCACCACCACCATCACGTGGATGACGACGACAAGATGAGCACCGCAAAACTG  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGGCACATCACCACCACCATCACGTGGATGACGACGACAAGATGAGCACCGCAAAACTG  60

Query  121  GTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGATGTGAGCGATAGCGaaaaa  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGATGTGAGCGATAGCGAAAAA  120

Query  181  aaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTCATCGATCTGAGCCTGATC  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  AAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTCATCGATCTGAGCCTGATC  180

Query  241  ACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATCGCAGTGCATGAAATGCTG  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  ACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATCGCAGTGCATGAAATGCTG  240

Query  301  GATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATGGCAGAACGTGCAGTGCAG  360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATGGCAGAACGTGCAGTGCAG  300

Query  361  CTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGCAAAACCCCGCTGAAAAGC  420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  CTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGCAAAACCCCGCTGAAAAGC  360

Query  421  TACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAACTGGCAGATCGTTACGCA  480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  TACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAACTGGCAGATCGTTACGCA  420

Query  481  ATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAAGATGATGATACCGCAGAT  540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  ATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAAGATGATGATACCGCAGAT  480

Query  541  ATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGGTTCATCGAATGCAACCTG  600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  481  ATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGGTTCATCGAATGCAACCTG  540

Query  601  GATCTGATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGT  660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  541  GATCTGATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGT  600

Query  661  TGA  663
            |||
Sbjct  601  TGA  603
```

Sequence Appendix 4 - Sequence alignment for all Bfr mutants used in chapter 4. Grey indicates the added or mutated codons. 'Query', obtained from sequence results from Supranom, Singapore, aligned against expected sequences 'Subject'. Aligned using Blast[10].

Designs and controls in pET-22b for lysate analysis

BfrWT

```
Query  52   ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  111
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  60

Query  112  GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  171
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  120

Query  172  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  231
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  180

Query  232  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  291
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  240

Query  292  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  351
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  300

Query  352  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  411
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  360

Query  412  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  471
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  420

Query  472  ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGTTAA  528
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGTTAA  477
```

BfrCC-N-term

```
Query  50   ATGTGTTGCAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAA  109
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGTGTTGCAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAA  60

Query  110  CTGGTGGCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAA  169
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   CTGGTGGCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAA  120

Query  170  CGTCTGAACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTAC  229
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  CGTCTGAACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTAC  180

Query  230  ATCGAACGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAAC  289
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  ATCGAACGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAAC  240

Query  290  ATCGGTGAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCA  349
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  ATCGGTGAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCA  300

Query  350  AAAAACCTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGAT  409
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAAACCTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGAT  360

Query  410  ATGATGATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTG  469
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  ATGATGATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTG  420

Query  470  GATCTGATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGT  529
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATCTGATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGT  480

Query  530  TAA  533
            |||
Sbjct  481  TAA  484
```

BfrCCPGCC-N-term

```
Query  49   ATGTGTTGTCCGGGCTGTTGTAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTG  108
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGTGTTGTCCGGGCTGTTGTAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTG  60

Query  109  CTGGGTAACGAACTGGTGGCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAAC  168
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   CTGGGTAACGAACTGGTGGCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAAC  120

Query  169  TGGGGTCTGAAACGTCTGAACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACAT  228
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  TGGGGTCTGAAACGTCTGAACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACAT  180

Query  229  GCAGATCGTTACATCGAACGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTG  288
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGATCGTTACATCGAACGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTG  240

Query  289  GGTAAACTGAACATCGGTGAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAA  348
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GGTAAACTGAACATCGGTGAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAA  300

Query  349  CTGGATGGTGCAAAAAACCTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTAC  408
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  CTGGATGGTGCAAAAAACCTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTAC  360

Query  409  GTGAGCCGTGATATGATGATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTG  468
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  GTGAGCCGTGATATGATGATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTG  420

Query  469  GAAACCGAACTGGATCTGATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATC  528
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GAAACCGAACTGGATCTGATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATC  480

Query  529  CGTGAAGAAGGTTAA  543
            |||||||||||||||
Sbjct  481  CGTGAAGAAGGTTAA  495
```

BfrCC-C-term

```
Query  51   ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  110
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  60

Query  111  GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  170
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  120

Query  171  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  230
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  180

Query  231  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  290
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Sbjct  181  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  240

Query  291  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  350
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  300

Query  351  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  410
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  360

Query  411  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  470
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  420

Query  471  ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGTTGCTGC  530
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGTTGCTGC  480

Query  531  TGA  534
            |||
Sbjct  481  TGA  484
```

BfrCC-N/C-term

```
Query  48   ATGTGTTGCAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAA  107
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGTGTTGCAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAA  60

Query  108  CTGGTGGCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAA  167
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   CTGGTGGCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAA  120

Query  168  CGTCTGAACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTAC  227
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  CGTCTGAACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTAC  180

Query  228  ATCGAACGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAAC  287
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  ATCGAACGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAAC  240

Query  288  ATCGGTGAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCA  347
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  ATCGGTGAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCA  300

Query  348  AAAAACCTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGAT  407
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAAACCTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGAT  360

Query  408  ATGATGATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTG  467
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  ATGATGATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTG  420

Query  468  GATCTGATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGT  527
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATCTGATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGT  480

Query  528  TGCTGCTGA  536
            |||||||||
Sbjct  481  TGCTGCTGA  489
```

BfrCCGPCC-C-term

```
Query  46   ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  105
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  60

Query  106  GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  165
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  120
```

```
Query  166  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  225
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  180

Query  226  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  285
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  240

Query  286  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  345
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  300

Query  346  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  405
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  360

Query  406  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  465
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  420

Query  466  ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGTTGCTGC  525
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGTTGCTGC  480

Query  526  GGCCCGTGCTGCTGA  540
            |||||||||||||||
Sbjct  481  GGCCCGTGCTGCTGA  495
```

BfrExt-C

```
Query  59   ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  118
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  60

Query  119  GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  178
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  120

Query  179  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  238
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  180

Query  239  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  298
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  240

Query  299  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  358
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  300

Query  359  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  418
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  360

Query  419  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  478
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  420

Query  479  ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGTTGTTGA  537
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGTTGTTGA  480
```

BfrExt-G-C

```
Query  48   ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  107
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  60
```

```
Query  108  GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  167
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  120

Query  168  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  227
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  180

Query  228  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  287
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  240

Query  288  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  347
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  300

Query  348  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  407
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  360

Query  408  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  467
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  420

Query  468  ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGTGGTTGT  527
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGTGGTTGT  480

Query  528  TGA  531
            |||
Sbjct  481  TGA  484
```

BfrN148C

```
Query  52   ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  111
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  60

Query  112  GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  171
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  120

Query  172  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  231
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  180

Query  232  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  291
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  240

Query  292  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  351
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  300

Query  352  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  411
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  360

Query  412  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  471
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  420

Query  472  ATCCAGAAAATGGGTCTGCAGTGCTACCTGCAGGCACAGATCCGTGAAGAAGGTTAA  528
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  ATCCAGAAAATGGGTCTGCAGTGCTACCTGCAGGCACAGATCCGTGAAGAAGGTTAA  477
```

BfrM147A,N148C

```
Query  49   ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  108
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  60

Query  109  GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  168
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  120

Query  169  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  228
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  180

Query  229  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  288
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  240

Query  289  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  348
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  300

Query  349  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  408
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  360

Query  409  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  468
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  420

Query  469  ATCCAGAAAATGGGTCTGGCTTGCTACCTGCAGGCACAGATCCGTGAAGAAGGTTAA  525
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  ATCCAGAAAATGGGTCTGGCTTGCTACCTGCAGGCACAGATCCGTGAAGAAGGTTAA  477
```

BFRQ151C

```
Query  63   ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  122
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  60

Query  123  GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  182
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  120

Query  183  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  242
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  180

Query  243  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  302
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  240

Query  303  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  362
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  300

Query  363  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  422
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  360

Query  423  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  482
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  420

Query  483  ATCCAGAAAATGGGTCTGCAGAACTACCTGTGCGCACAGATCCGTGAAGAAGGTTAA  539
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  ATCCAGAAAATGGGTCTGCAGAACTACCTGTGCGCACAGATCCGTGAAGAAGGTTAA  477
```

BfrE157C

```
Query  59   ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  118
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Sbjct  1    ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  60

Query  119  GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  178
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  120

Query  179  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  238
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  180

Query  239  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  298
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  240

Query  299  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  358
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  300

Query  359  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  418
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  360

Query  419  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  478
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  420

Query  479  ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAATGTGGTTAA  535
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAATGTGGTTAA  477
```

BfrCC-C-termN23F

```
Query  64   ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  123
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  60

Query  124  GCAATCTTCCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  183
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GCAATCTTCCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  120

Query  184  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  243
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  180

Query  244  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  303
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  240

Query  304  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  363
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  300

Query  364  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  423
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  360

Query  424  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  483
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  420

Query  484  ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGTTGCTGC  543
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGTTGCTGC  480

Query  544  TGA  547
             |||
Sbjct  481  TGA  484
```

BfrCC-C-termD118F

```
Query  63   ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  122
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  60

Query  123  GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  182
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  120

Query  183  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  242
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  180

Query  243  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  302
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  240

Query  303  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  362
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  300

Query  363  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTTTCATGATG  422
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTTTCATGATG  360

Query  423  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  482
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  420

Query  483  ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGTTGCTGC  542
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGTTGCTGC  480

Query  543  TGA  546
            |||
Sbjct  481  TGA  484
```

BFRQ151C,N23F

```
Query  53   ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  112
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  60

Query  113  GCAATCTTCCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  172
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GCAATCTTCCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  120

Query  173  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  232
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  180

Query  233  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  292
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  240

Query  293  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  352
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  300

Query  353  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  412
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  360

Query  413  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  472
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  420

Query  473  ATCCAGAAAATGGGTCTGCAGAACTACCTGTGCGCACAGATCCGTGAAGAAGGTTGA  529
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  ATCCAGAAAATGGGTCTGCAGAACTACCTGTGCGCACAGATCCGTGAAGAAGGTTGA  477
```

BFRQ151C,D118F

```
Query  48   ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  107
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  60

Query  108  GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  167
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  120

Query  168  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  227
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  180

Query  228  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  287
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  240

Query  288  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  347
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  300

Query  348  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTTTCATGATG  407
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTTTCATGATG  360

Query  408  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  467
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  420

Query  468  ATCCAGAAAATGGGTCTGCAGAACTACCTGTGCGCACAGATCCGTGAAGAAGGTTGA  524
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  ATCCAGAAAATGGGTCTGCAGAACTACCTGTGCGCACAGATCCGTGAAGAAGGTTGA  477
```

BfrE157C,N23F

```
Query  51   ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  110
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  60

Query  111  GCAATCTTCCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  170
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GCAATCTTCCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  120

Query  171  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  230
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  180

Query  231  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  290
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  240

Query  291  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  350
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  300

Query  351  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  410
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  360

Query  411  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  470
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  420

Query  471  ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAATGTGGTTGA  527
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Sbjct  421  ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAATGTGGTTGA  477
```

BfrE157C,D118F

```
Query  49   ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  108
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  60

Query  109  GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  168
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  120

Query  169  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  228
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  180

Query  229  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  288
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  240

Query  289  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  348
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  300

Query  349  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTTTCATGATG  408
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTTTCATGATG  360

Query  409  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  468
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  420

Query  469  ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAATGTGGTTGA  525
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAATGTGGTTGA  477
```

Designs and controls in pET-32b to allow for protein purification

BfrWT

```
Query  543   ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  602
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1     ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  60

Query  603   GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  662
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61    GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  120

Query  663   AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  722
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121   AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  180

Query  723   CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  782
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181   CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  240

Query  783   GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  842
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241   GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  300

Query  843   CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  902
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301   CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  360

Query  903   ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  962
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361   ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  420

Query  963   ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGTTGA  1019
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421   ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGTTGA  477
```

BfrCC-C-term

```
Query  535   ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  594
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1     ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  60

Query  595   GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  654
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61    GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  120

Query  655   AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  714
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121   AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  180

Query  715   CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  774
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181   CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  240

Query  775   GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  834
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241   GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  300

Query  835   CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  894
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301   CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  360

Query  895   ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  954
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361   ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  420

Query  955   ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGTTGCTGC
1014
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421   ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGTTGCTGC  480

Query  1015  TAA  1028
              |||
Sbjct  481   TAA  484
```

BfrCC-N/C-term

```
Query  537   ATGTGCTGCAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAA  596
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1     ATGTGCTGCAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAA  60

Query  597   CTGGTGGCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAA  656
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61    CTGGTGGCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAA  120

Query  657   CGTCTGAACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTAC  716
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121   CGTCTGAACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTAC  180

Query  717   ATCGAACGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAAC  776
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181   ATCGAACGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAAC  240

Query  777   ATCGGTGAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCA  836
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241   ATCGGTGAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCA  300

Query  837   AAAAACCTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGAT  896
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301   AAAAACCTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGAT  360

Query  897   ATGATGATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTG  956
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  ATGATGATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTG  420

Query  957  GATCTGATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGT
1016
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATCTGATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGT  480

Query  1017 TGCTGCTAA  1025
            |||||||||
Sbjct  481  TGCTGCTAA  489
```

BfrCCPGCC-N-term

```
Query  527  ATGTGTTGTCCGGGCTGTTGTAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTG  586
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGTGTTGTCCGGGCTGTTGTAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTG  60

Query  587  CTGGGTAACGAACTGGTGGCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAAC  646
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   CTGGGTAACGAACTGGTGGCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAAC  120

Query  647  TGGGGTCTGAAACGTCTGAACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACAT  706
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  TGGGGTCTGAAACGTCTGAACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACAT  180

Query  707  GCAGATCGTTACATCGAACGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTG  766
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGATCGTTACATCGAACGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTG  240

Query  767  GGTAAACTGAACATCGGTGAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAA  826
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GGTAAACTGAACATCGGTGAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAA  300

Query  827  CTGGATGGTGCAAAAAACCTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTAC  886
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  CTGGATGGTGCAAAAAACCTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTAC  360

Query  887  GTGAGCCGTGATATGATGATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTG  946
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  GTGAGCCGTGATATGATGATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTG  420

Query  947  GAAACCGAACTGGATCTGATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATC
1006
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GAAACCGAACTGGATCTGATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATC  480

Query  1007 CGTGAAGAAGGTTAA  1021
            |||||||||||||||
Sbjct  481  CGTGAAGAAGGTTAA  495
```

BfrCCGPCC-C-term

```
Query  528  ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  587
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  60

Query  588  GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  647
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  120

Query  648  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  707
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  180

Query  708  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  767
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  240
```

```
Query  768   GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  827
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241   GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  300

Query  828   CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  887
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301   CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  360

Query  888   ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  947
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361   ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  420

Query  948   ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGTTGCTGC
1007
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421   ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGTTGCTGC  480

Query  1008  GGCCCGTGCTGCTAA  1022
              |||||||||||||||
Sbjct  481   GGCCCGTGCTGCTAA  495
```

BfrCC-C-term,N23F

```
Query  42    ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  101
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1     ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  60

Query  102   GCAATCTTCCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  161
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61    GCAATCTTCCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  120

Query  162   AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  221
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121   AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  180

Query  222   CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  281
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181   CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  240

Query  282   GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  341
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241   GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  300

Query  342   CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  401
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301   CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  360

Query  402   ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  461
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361   ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  420

Query  462   ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGTTGCTGC  521
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421   ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGTTGCTGC  480

Query  522   TGA  524
              |||
Sbjct  481   TGA  483
```

BfrCC-C-term,D118F

```
Query  26    ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  85
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1     ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  60

Query  86    GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  145
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61    GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  120
```

```
Query  146  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  205
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  180

Query  206  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  265
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  240

Query  266  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  325
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  300

Query  326  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTTTCATGATG  385
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTTTCATGATG  360

Query  386  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  445
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  420

Query  446  ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGTTGCTGC  505
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGTTGCTGC  480

Query  506  TGA  508
            |||
Sbjct  481  TGA  483
```

BfrCCGPCC-C-term,N23F

```
Query  30   ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  89
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  60

Query  90   GCAATCTTCCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  149
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GCAATCTTCCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  120

Query  150  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  209
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  180

Query  210  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  269
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  240

Query  270  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  329
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  300

Query  330  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  389
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTGATATGATG  360

Query  390  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  449
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  420

Query  450  ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGTTGCTGC  509
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGTTGCTGC  480

Query  510  GGCCCGTGCTGCTGA  524
            |||||||||||||||
Sbjct  481  GGCCCGTGCTGCTGA  495
```

BfrCCGPCC-C-term,D118F

```
Query  27   ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  86
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAAAGGTGATACCAAAGTGATCAACTACCTGAACAAACTGCTGGGTAACGAACTGGTG  60

Query  87   GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  146
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GCAATCAACCAGTACTTCCTGCATGCACGTATGTTCAAAAACTGGGGTCTGAAACGTCTG  120

Query  147  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  206
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  AACGATGTGGAATACCATGAAAGCATCGATGAAATGAAACATGCAGATCGTTACATCGAA  180

Query  207  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  266
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  CGTATCCTGTTCCTGGAAGGTCTGCCGAACCTGCAGGATCTGGGTAAACTGAACATCGGT  240

Query  267  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  326
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GAAGATGTGGAAGAAATGCTGCGTAGCGATCTGGCACTGGAACTGGATGGTGCAAAAAAC  300

Query  327  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTTTCATGATG  386
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  CTGCGTGAAGCAATCGGTTACGCAGATAGCGTGCATGATTACGTGAGCCGTTTCATGATG  360

Query  387  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  446
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  ATCGAAATCCTGCGTGATGAAGAAGGTCATATCGATTGGCTGGAAACCGAACTGGATCTG  420

Query  447  ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGTTGCTGC  506
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  ATCCAGAAAATGGGTCTGCAGAACTACCTGCAGGCACAGATCCGTGAAGAAGGTTGCTGC  480

Query  507  GGCCCGTGCTGCTGA  521
            |||||||||||||||
Sbjct  481  GGCCCGTGCTGCTGA  495
```

Sequence Appendix 5 - Sequence alignment for all Dps mutants used in chapter 5. Grey indicates the added or mutated codons. 'Query', obtained from sequence results from Eurofins, Germany, aligned against expected sequences 'Subject'. Aligned using Blast[10].

Designs and controls in pET-22b for lysate analysis

DpsR18C,N19C,A126C,N130C

```
Query  51   ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCTGCTGCGAT  110
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCTGCTGCGAT  60

Query  111  GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  170
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  171  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  230
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  231  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  290
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  291  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  350
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  351  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  410
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  411  CTGGCAGATCGTTACTGCATCGTGGCATGCGATGTGCGTAAAGCAATCGGTGAAGCAAAA  470
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCAGATCGTTACTGCATCGTGGCATGCGATGTGCGTAAAGCAATCGGTGAAGCAAAA  420

Query  471  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  530
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  531  TTCATCGAATGCAACATCGAATAA  554
            ||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAATAA  504
```

DpsR18C,N19C,D123C,A126C

```
Query  62   ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCTGCTGCGAT  121
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCTGCTGCGAT  60

Query  122  GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  181
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  182  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  241
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  242  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  301
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  302  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  361
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  362  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  421
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  422  CTGGCATGCCGTTACTGCATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  481
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCATGCCGTTACTGCATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  420

Query  482  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  541
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  542  TTCATCGAATGCAACATCGAATAA  565
            |||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAATAA  504
```

DpsN19C,D20C,D123C,A126C

```
Query  60   ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTTGCTGC  119
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTTGCTGC  60

Query  120  GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  179
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  180  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  239
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  240  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  299
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  300  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  359
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  360  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  419
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  420  CTGGCATGCCGTTACTGCATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  479
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCATGCCGTTACTGCATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  420

Query  480  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  539
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  540  TTCATCGAATGCAACATCGAATAA  563
            |||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAATAA 504
```

DpsN19C,D20C,A126C,N130C

```
Query  59   ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTTGCTGC  118
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTTGCTGC  60

Query  119  GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  178
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  179  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  238
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  239  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  298
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  299  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  358
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  359  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  418
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  419  CTGGCAGATCGTTACTGCATCGTGGCATGCGATGTGCGTAAAGCAATCGGTGAAGCAAAA  478
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCAGATCGTTACTGCATCGTGGCATGCGATGTGCGTAAAGCAATCGGTGAAGCAAAA  420

Query  479  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  538
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  539  TTCATCGAATGCAACATCGAATAA  562
            ||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAATAA  504
```

DpsR18C,D20C,D123C,A126C,N130C

```
Query  60   ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCTGCAACTGC  119
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCTGCAACTGC  60

Query  120  GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  179
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  180  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  239
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  240  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  299
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  300  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  359
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  360  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  419
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  420  CTGGCATGCCGTTACTGCATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  479
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCATGCCGTTACTGCATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  420

Query  480  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  539
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  540  TTCATCGAATGCAACATCGAATAA  563
            ||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAATAA  504
```

DpsR18C,D20C,A126C,N130C

```
Query  57   ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCTGCAACTGC  116
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCTGCAACTGC  60

Query  117  GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  176
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120
```

```
Query  177  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  236
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  237  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  296
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  297  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  356
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  357  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  416
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  417  CTGGCAGATCGTTACTGCATCGTGGCATGCGATGTGCGTAAAGCAATCGGTGAAGCAAAA  476
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCAGATCGTTACTGCATCGTGGCATGCGATGTGCGTAAAGCAATCGGTGAAGCAAAA  420

Query  477  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  536
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  537  TTCATCGAATGCAACATCGAATAA  560
            ||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAATAA  504
```

DpsWT

```
Query  50   ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  109
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  60

Query  110  GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  169
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  170  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  229
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  230  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  289
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  290  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  349
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  350  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  409
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  410  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  469
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  420

Query  470  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  529
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  530  TTCATCGAATGCAACATCGAATAA  553
            ||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAATAA  504
```

DpsCCPGCC

```
Query  41   ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  60

Query  101  GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  161  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  221  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  281  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  341  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  400
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  401  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  460
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  420

Query  461  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  520
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  521  TTCATCGAATGCAACATCGAATGCTGCCCAGGTTGCTGCTAA  562
            ||||||||||||||||||||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAATGCTGCCCAGGTTGCTGCTAA  522
```

DpsCC

```
Query  30   ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  89
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  60

Query  90   GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  149
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  150  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  209
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  210  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  269
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  270  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  329
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  330  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  389
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  390  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  449
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  420

Query  450  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  509
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  510  TTCATCGAATGCAACATCGAATGCTGCTGA  539
            ||||||||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAATGCTGCTGA  510
```

DpsGCC

```
Query  14   ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  73
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  60

Query  74   GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  133
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  134  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  193
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  194  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  253
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  254  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  313
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  314  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  373
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  374  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  433
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  420

Query  434  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  493
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  494  TTCATCGAATGCAACATCGAAGGATGCTGCTGA  526
            ||||||||||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAAGGATGCTGCTGA  513
```

DpsGGCC

```
Query  40   ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  99
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  60

Query  100  GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  159
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  160  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  219
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  220  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  279
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  280  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  339
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  340  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  399
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  400  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  459
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  420

Query  460  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  519
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  520  TTCATCGAATGCAACATCGAAGGAGGTTGCTGCTGA  555
            ||||||||||||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAAGGAGGTTGCTGCTGA  516
```

DpsGGGCC

```
Query  37   ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  96
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  60

Query  97   GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  156
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  157  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  216
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  217  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  276
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  277  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  336
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  337  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  396
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  397  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  456
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  420

Query  457  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  516
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  517  TTCATCGAATGCAACATCGAAGGAGGTGGATGCTGCTGA  555
            ||||||||||||||||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAAGGAGGTGGATGCTGCTGA  519
```

DpsPAGCC

```
Query  39   ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  98
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  60

Query  99   GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  158
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  159  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  218
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  219  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  278
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240
```

```
Query  279  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  338
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  339  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  398
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  399  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  458
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  420

Query  459  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  518
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  519  TTCATCGAATGCAACATCGAACCAGCGGGATGTTGCTGA  557
            |||||||||||||||||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAACCAGCGGGATGTTGCTGA  519
```

Designs and controls in pET-32b to allow for protein purification

DpsWT

```
Query  9    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  68
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  60

Query  69   GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  128
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  129  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  188
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  189  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  248
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  249  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  308
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  309  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  368
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  369  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  428
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  420

Query  429  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  488
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  489  TTCATCGAATGCAACATCGAATAA  512
            |||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAATAA  504
```

DpsCCPGCC

```
Query  9    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  68
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  60

Query  69   GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  128
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  129  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  188
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  189  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  248
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  249  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  308
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  309  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  368
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  369  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  428
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  420

Query  429  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  488
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  489  TTCATCGAATGCAACATCGAATGCTGCCCAGGTTGCTGCTAA  530
            ||||||||||||||||||||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAATGCTGCCCAGGTTGCTGCTAA  522
```

DpsGGGCC

```
Query  9    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  68
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  60

Query  69   GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  128
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  129  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  188
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  189  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  248
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  249  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  308
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  309  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  368
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  369  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  428
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  420

Query  429  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  488
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  489  TTCATCGAATGCAACATCGAAGGAGGTGGATGCTGCTGA  527
            ||||||||||||||||||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAAGGAGGTGGATGCTGCTGA  519
```

DpsPAGCC

```
Query  11   ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  70
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  60

Query  71   GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  130
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  131  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  190
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  191  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  250
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  251  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  310
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  311  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  370
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  371  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  430
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  420

Query  431  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  490
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  491  TTCATCGAATGCAACATCGAACCAGCGGGATGTTGCTGA  529
            |||||||||||||||||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAACCAGCGGGATGTTGCTGA  519
```

DpsR18C,N19C,D123C,A126C

```
Query  534   ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCTGCTGCGAT  593
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1     ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCTGCTGCGAT  60

Query  594   GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  653
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61    GTGAGCGATAGCGAAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  654   ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  713
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121   ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  714   GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  773
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181   GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  774   GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  833
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241   GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  834   AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  893
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301   AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  894   CTGGCATGCCGTTACTGCATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  953
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361   CTGGCATGCCGTTACTGCATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  420
```

```
Query  954   GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG
1013
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421   GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  1014  TTCATCGAATGCAACATCGAATAA  1037
              ||||||||||||||||||||||||
Sbjct  481   TTCATCGAATGCAACATCGAATAA  504
```

Sequence Appendix 6 - Sequence alignment for all Dps mutants used in chapter 6. Grey indicates the added or mutated codons. 'Query', obtained from sequence results from Eurofins, Germany, aligned against expected sequences 'Subject'. Aligned using Blast[10].

Designs and controls in pET-46 for both *in vitro* and *in vivo* analysis

DpsCCPGCC

```
Query  85   ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  144
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  60

Query  145  GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  204
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  205  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  264
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  265  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  324
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  325  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  384
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  385  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  444
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  445  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  504
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  420

Query  505  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  564
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  565  TTCATCGAATGCAACATCGAATGCTGCCCAGGTTGCTGCTAA  606
            ||||||||||||||||||||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAATGCTGCCCAGGTTGCTGCTAA  522
```

DpsPAGCC

```
Query  84   ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  143
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  60

Query  144  GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  203
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  204  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  263
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  264  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  323
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  324  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  383
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  384  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  443
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  444  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  503
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  420

Query  504  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  563
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  564  TTCATCGAATGCAACATCGAACCAGCGGGATGTTGCTGA  602
            ||||||||||||||||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAACCAGCGGGATGTTGCTGA  519
```

DpsWT

```
Query  79   ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  138
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  60

Query  139  GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  198
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  199  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  258
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  259  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  318
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  319  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  378
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  379  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  438
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  439  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  498
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  420

Query  499  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  558
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  559  TTCATCGAATGCAACATCGAATAA  582
            ||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAATAA  504
```

Dps(AA)PAGCC

```
Query  100  ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  159
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  60

Query  160  GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  219
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  220  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  279
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  280  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  339
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  340  GCAGAAGCAGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  399
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAAGCAGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  400  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  459
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  460  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGGCAAAAGCAATCGGTGAAGCAAAA  519
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGGCAAAAGCAATCGGTGAAGCAAAA  420

Query  520  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  579
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  580  TTCATCGAATGCAACATCGAACCAGCGGGATGTTGCTGA  618
             ||||||||||||||||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAACCAGCGGGATGTTGCTGA  519
```

Dps(DD)PAGCC

```
Query  110  ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  169
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  60

Query  170  GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  229
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  230  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  289
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  290  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  349
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  350  GCAGAAGACGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  409
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAAGACGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  410  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  469
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  470  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGGACAAAGCAATCGGTGAAGCAAAA  529
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGGACAAAGCAATCGGTGAAGCAAAA  420

Query  530  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  589
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  590  TTCATCGAATGCAACATCGAACCAGCGGGATGTTGCTGA  628
             ||||||||||||||||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAACCAGCGGGATGTTGCTGA  519
```

Dps(WW)PAGCC

```
Query  70   ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  129
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  60

Query  130  GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  189
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120
```

```
Query  190  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  249
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  250  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  309
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  310  GCAGAATGGGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  369
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAATGGGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  370  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  429
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  430  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGTGGAAAGCAATCGGTGAAGCAAAA  489
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGTGGAAAGCAATCGGTGAAGCAAAA  420

Query  490  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  549
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  550  TTCATCGAATGCAACATCGAACCAGCGGGATGTTGCTGA  588
            ||||||||||||||||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAACCAGCGGGATGTTGCTGA  519
```

Dps(AA)PAGCC^{Q86V,L87F,N130F}

```
Query  99   ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  158
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  60

Query  159  GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  218
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  219  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  278
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  279  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  338
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  339  GCAGAAGCAGCAGTGGTCTTCGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  398
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAAGCAGCAGTGGTCTTCGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  399  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  458
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  459  CTGGCAGATCGTTACGCAATCGTGGCATTCGATGTGGCAAAAGCAATCGGTGAAGCAAAA  518
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCAGATCGTTACGCAATCGTGGCATTCGATGTGGCAAAAGCAATCGGTGAAGCAAAA  420

Query  519  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  578
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  579  TTCATCGAATGCAACATCGAACCAGCGGGATGTTGCTGA  617
            ||||||||||||||||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAACCAGCGGGATGTTGCTGA  519
```

Dps(AA)PAGCC^{N130Y}

```
Query  94   ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  153
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  60

Query  154  GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  213
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  214  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  273
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  274  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  333
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  334  GCAGAAGCAGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  393
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAAGCAGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  394  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  453
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  454  CTGGCAGATCGTTACGCAATCGTGGCATACGATGTGGCAAAAGCAATCGGTGAAGCAAAA  513
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCAGATCGTTACGCAATCGTGGCATACGATGTGGCAAAAGCAATCGGTGAAGCAAAA  420

Query  514  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  573
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  574  TTCATCGAATGCAACATCGAACCAGCGGGATGTTGCTGA  612
            ||||||||||||||||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAACCAGCGGGATGTTGCTGA  519
```

Dps(AA)PAGCC [N130L,D141D,D142W,D143G]

```
Query  111  ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  170
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  60

Query  171  GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  230
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  231  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  290
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  291  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  350
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  351  GCAGAAGCAGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  410
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAAGCAGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  411  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  470
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  471  CTGGCAGATCGTTACGCAATCGTGGCACTCGATGTGGCAAAAGCAATCGGTGAAGCAAAA  530
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCAGATCGTTACGCAATCGTGGCACTCGATGTGGCAAAAGCAATCGGTGAAGCAAAA  420

Query  531  GACTGGGGCACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  590
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GACTGGGGCACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480
```

```
Query  591  TTCATCGAATGCAACATCGAACCAGCGGGATGTTGCTGA  629
            |||||||||||||||||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAACCAGCGGGATGTTGCTGA  519
```

References

1. Grant, R. A.; Filman, D. J.; Finkel, S. E.; Kolter, R.; Hogle, J. M., The crystal structure of Dps, a ferritin homolog that binds and protects DNA. *Nature Structural Biology* 1998, *5* (4), 294-303.
2. Bryce, C. F. A.; Crichton, R. R., Catalytic activity of horse spleen apoferritin - preliminary kinetic studies and effect of chemical modification. *Biochemical Journal* 1973, *133* (2), 301-&.
3. Dautant, A.; Meyer, J. B.; Yariv, J.; Precigoux, G.; Sweet, R. M.; Kalb, A. J.; Frolow, F., Structure of a monoclinic crystal form of cytochrome b1 (bacterioferritin) from E-coli. *Acta Crystallographica Section D-Biological Crystallography* 1998, *54*, 16-24.
4. Fan, R.; Boyle, A. L.; Cheong, V. V.; Ng, S. L.; Orner, B. P., A Helix Swapping Study of Two Protein Cages. *Biochemistry* 2009, *48* (24), 5623-5630.
5. Zhang, Y.; Fu, J.; Chee, S. Y.; Ang, E. X. W.; Orner, B. P., Rational disruption of the oligomerization of the mini-ferritin E. coli DPS through protein-protein interface mutation. *Protein Science* 2011, *20* (11), 1907-1917.
6. Zhang, Y.; Orner, B. P., Self-Assembly in the Ferritin Nano-Cage Protein Superfamily. *International Journal of Molecular Sciences* 2011, *12* (8), 5406-5421.
7. Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W., NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* 2012, *9* (7), 671-675.
8. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* 2004, *25* (13), 1605-1612.
9. Sobolev, V.; Eyal, E.; Gerzon, S.; Potapov, V.; Babor, M.; Prilusky, J.; Edelman, M., SPACE: a suite of tools for protein structure prediction and analysis based on complementarity and environment. *Nucleic Acids Research* 2005, *33*, W39-W43.
10. Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J., Basic Local Alignment Search Tool. *Journal of Molecular Biology* 1990, *215* (3), 403-410.

Publish work 1

Cover page

Authors – Thomas A. Cornell[1] and Brendan P. Orner[1]

[1] King's College London, Department of Chemistry, Guys Campus, London, SE1 1DB

Thomas.cornell@kcl.ac.uk

Brendan.orner@kcl.ac.uk

Detection of Protein Cage Assembly through Probing with BisArsenic Fluorophores.

Thomas A. Cornell and Brendan P. Orner

Summary

We describe a method for the detection of specific protein-protein interactions in protein cages through the exploitation of designed binding sites for bisarsinic fluorescent probes. These sites are engineered to be protein-protein interface specific. We have adapted this method to ferritins, however it could conceivably be applied to other protein cages. It is thought that this technique could be utilised in the thermodynamic and kinetic characterisation of cage assembly mechanisms and in the high throughput screening of protein cage libraries for the discovery of proteins with new assembly properties or of optimised conditions for assembly.

Keywords – FlAsH-EDT_2, Protein-protein interactions, Fluorescence detection, Oligomerization state, Ferritin.

Running head – Fluorescence assay to detect protein cage formation.

1. Introduction

Nanocage proteins play key roles across many biological systems. They perform wide-ranging functions, such as metabolite synthesis (*1*), protein folding assistance (*2*), viral genome protection and delivery (*3*), and ion and metabolite storage and sequestration (*4*), which are dependent on, and a consequence of, their large, complex, and often self-

assembly controlled, structures (*5*). Many of the techniques employed to study the formation of these unique architectures are often indirect and provide only low structural and temporal resolution.

Because of their unique hollow nanostructure, protein cages are attractive for a wide range of non-natural applications. For example, they have been employed as vehicles capable of controlled drug delivery and as size-constrained synthetic reaction vessels for mineralised materials (*6, 7*). To develop such systems with greater utility, proteins with properties tailored to each specific application are required. However, the rational design of protein cages with bespoke properties, faces the dual challenges of engineering both protein folding and protein-protein interactions. While advancing (*6, 8, 9*), these both are far from solved problems. Thus, it is envisioned that high throughput techniques to discover protein cages with novel properties would not only be helpful but are necessary.

The complex and often highly symmetric structures of many protein cages are stitched together by protein-protein interactions between their protein building blocks. Understanding the fundamentals controlling the assembly could help scientists to utilise better and to engineer protein cages in the future. Research to dissect the role of the individual protein-protein interactions and their components can be slow and tedious, utilising often limited and indirect techniques, such as alanine scanning via site directed mutagenesis, followed by size exclusion chromatography (SEC) and transmission electron microscopy (TEM) for characterisation of each of the mutants individually (*10, 11, 12, 13, 14, 15*). While these strategies have provided much insight, a more rapid method could enhance the field by acquiring this information more swiftly. Furthermore, a technique that more directly assesses specific protein-protein interactions could provide greater utility in this field.

A growing number of methods, such as Förster resonance energy transfer (FRET),split green fluorescent proteins (GFP) and split luciferase enzymes (*16, 17, 18, 19*), have been employed to study the assembly of two protein binding partners. Such techniques often require very large protein fusions (*20*), antibodies (*21*), or post-translational modifications (*22, 23*) and are often limited to two binding partners.

Recently bisarsenic fluorescent reagents, such as FlAsH and ReAsH, have been developed as protein probes. These reagents become fluorescent when bound to a protein with four cysteine residues displayed in a specific geometric relationship (*24*). The reagents provide a smaller labelling tag than GFP fusions and thus allow the probing of proteins in a more native state within cells (*25*). Moreover, by splitting the tetra cysteine peptide tag into two cysteine pairs that straddle a protein-protein interface, it is possible to design binding sites for these reagents that are dependent upon the formation of specific protein-protein interactions (*26, 27*). While this method was originally shown to work for short peptide chains, more recently it has been used to establish structurally detailed insight into the conformationally driven transduction of binding information in the trans-membrane helix of EGFR (*28*).

We have recently expanded this technique to interrogate the self-assembly of a large multimeric ferritin protein cage (see Figure 1) (*29*). In our study, the ferritin protein, DNA-binding protein from starved cells (Dps), was engineered with a C-terminal bipartite cysteine pair. The C-termini of Dps monomers converge when this protein assembles into a tetrahedrally symmetric cage but diverge in the presumed two-fold symmetric dimer intermediate of the assembly process. The convergent cysteines create a binding site for the bisarsenic fluorescent reagent, resulting in a cage-dependent signal. We have demonstrated that this technique works in bacterial lysates thus removing the need to purify and analyse each protein mutant individually, greatly

increasing throughput. This flexible strategy could be used to establish the assembly dynamics of protein cages *in vitro* or in cells. In a high throughput format, it also has potential for the discovery of mutant protein cages with enhanced properties or for screening conditions for cage stability and assembly. The following chapter describes a method for the application of this technique.

[INSERT FIGURE 1 HERE]

2. Materials

2.1. Protein design and preparation

1. PDB coordinates for the protein cage.
2. Inducible expression plasmid containing gene encoding the protein cage and carrying antibiotic resistance.
3. Primers for site directed mutagenesis or extension PCR.
4. Expression host cell line such as Rosetta (BL21) *E. coli* (Novagen).
5. Luria Broth (LB): 10 g Tryptone plus (Sigma), 5 g Yeast extract (Sigma), 10 g NaCl (Sigma), 1 L water. Autoclave.
6. Antibiotics for inoculation (such as Carbenicillin, 50 mg/mL) and reagent to activate expression (such as IPTG, 0.5 mM final concentration).
7. FlAsH buffer: 100 mM Tris.HCl (Sigma), 100 mM NaCl (Sigma), 1 mM EDTA (Sigma), pH 7.8. Degas.

2.2. FlAsH analysis

1. FlAsH-EDT_2 (Invitrogen): 1 μM in FlAsH buffer.

2. Additives for signal and specificity optimisation: 1, 2-Ethanedithiol (EDT): 1 mM (Sigma), 2-Mercaptoethanol (2-ME): 1 mM (Sigma), *Tris*(2-carboxyethyl)phosphine (TCEP): 3.5 mM (Sigma). Make all stocks in FlAsH buffer immediately before use.

3. Guanidine-HCl (Sigma): 8 M in FlAsH buffer.

4. Steady state fluorimeter or fluorescent plate reader.

3. Methods

The following procedure should be used as a guide to help utilise the fluorescent reagent FlAsH-EDT_2 to detect specific protein-protein interactions in protein cages. Because these interactions arise only when the cages assemble into higher order oligomerization states, this strategy is based on designing FlAsH binding sites at the protein-protein interfaces that appear when the cage forms. The binding sites are generated when bipartite cysteine pairs engineered into the monomers become proximal upon cage assembly. Once the binding sites emerge, *in vitro* or *in vivo* detection of oligomerization is possible by observing fluorescence.

3.1. Designing binding sites dependent upon cage assembly

1. Design FlAsH binding sites guided by the inspection of a protein data bank (PDB) structure. Position two pairs of cysteines across an interface that ideally only forms in the highest order cage structure and does not appear in any assembly intermediate (unless the identification of such an intermediate is the goal of the study), (see **Note 1**, **2**, **3**, **4**, **5**, and see Figure 2).

[INSERT FIGURE 2 HERE]

2. Design a positive control protein by fusing a peptide sequence containing CCPGCC (there are several optimised versions) (*30*) onto a monomer. This is most easily done as an extension to one of the termini (see **Note 6**).

3.2. Protein preparation

1. Use standard extension or site directed mutagenesis cloning techniques to introduce mutations required to generate the designed proteins (see **Note 7**, **8**).

2. Express the designed proteins by first transferring constructed plasmids into an expression host (for example Rosetta, *E. coli* expression cells) and induce protein expression (such as by growing in inoculated LB until an $O.D_{600}$ of 0.6 followed by the addition of IPTG), (see **Note 9**).

3. Isolate cell pellet by centrifugation, disrupt cells (for example by sonication), clarify by centrifugation to obtain the soluble proteins (see **Note 10**).

4. If choosing to assay the proteins in purified form, as opposed to in lysates, isolate and characterise the proteins with size exclusion chromatography (SEC) (see **Note 11**, **12**, **13**, **14**) to confirm assembly.

3.4. Sample preparation and fluorescence analysis

Initial screening in lysate samples can be used to rapidly guide the design process. Samples can either be screened in a high throughput plate format, which speeds up data collection and mitigates the running of replicate experiments, or in cuvettes with a fluorimeter (See **Note 15**). Experiments should be designed to include a positive control protein, which not only can help establish that the dye is working as expected, but also can provide a detection upper limit to benchmark optimised dye and protein concentrations and the

sensitivity of the fluorimeter. A negative control, typically a wild-type protein with no additional cysteines, should also be included. This control determines the level of the background fluorescence due to non-selective binding to the protein. A no-protein control should also be used to establish further sources of background from the FlAsH itself.

1. Dilute the lysate sample to the required total protein concentration (0.1 mg/ml, BCA, Merck) with FlAsH buffer and add EDT, 2-ME and TCEP solutions and incubate for two hours at room temperature (see **Note 16**, **17**, **18** and see Figure 3).

[INSERT FIGURE 3 HERE]

2. Add FlAsH-EDT_2 solution and incubate for an additional two hours at room temperature away from light (see **Note 19 and 20**).

3. Read each sample in a fluorimeter or plate reader. Multiple replicates are suggested (see **Note 21**).

3.5. Denatured protein fluorescence analysis

A useful control for multimeric proteins is to assay the designs, but in the presence of a denaturant such as 6 M Guanidine-HCl. At high enough concentrations of the denaturant, the proteins should unfold and all quaternary structure will be lost. This will destroy the designed FlAsH binding sites, resulting in the loss of any fluorescent signal observed in assembly promoting conditions. This should confirm that the signal is indeed oligomerization independent especially if the positive control designs containing CCPGCC are not affected by the denaturant.

1. For protein denaturation experiments, dilute the sample to the required protein concentration with 8 M Guanidine-HCl in FlAsH buffer, until a final concentration of 6 M Guanidine-HCl is achieved. Incubate for two hours at room temperature.

2. Add EDT, 2-ME and TCEP solutions and incubate for a further two hours at room temperature (see above).

3. Add FlAsH-EDT_2 and incubate for 2 more hours, away from light at room temperature. Run each sample in a fluorimeter and compare to non-denatured samples (see **Note 22**).

4. Notes

1. For large multimeric nanostructures, look for interfaces that form only when the highest oligomerization state is achieved. By directing the design towards highly symmetric interfaces on protein nanostructures, it is possible to increase the number of binding sites per structure allowing for detection at lower protein concentrations.

2. The two-fold symmetry dimer is presumed to be an assembly intermediate in most mini-and maxi-ferritin cages. The protein-protein interface of this dimer is avoided for binding site engineering as it would exist regardless of the formation of higher order cage structures.

3. Termini that converge at interfaces make for ‘quick and easy’ locations for the addition of cysteine pairs, however by using this strategy, the potential for an optimal geometry may be unachievable, and may require flexible termini. Although some termini may be close enough in space to allow the cysteine pairs

to form an appropriate FlAsH binding site, others may require screening a range of linkers in order to optimise FlAsH binding (see Figure 2).

4. Binding site designs that employ a more structurally complex strategy than simply attaching cysteines to termini may initially be thought to be preferred. However, care must be taken when swapping one amino acid for a cysteine as this could lead to a) protein misfolding due to undesired disulphide bond formation or b) the removal of key hot-spots residues essential for stabilising protein-protein interactions. It is reasonable to pursue both strategies in parallel.

5. Take note of all native cysteines in the structure. Any that are too close to the FlAsH binding site might need to be mutated out; on the other hand it could be imagined that some native cysteines could be included in the engineered binding site.

6. This can be a more useful control than simply using the synthesized small peptide, because this protein can also be employed as a control for oligomerization/denaturation studies to determine how assembly-dependent the designed interfacial binding site is.

7. If the protein cage monomers can be expressed in a soluble form and assemble without extraordinary measures, then it is possible that the fluorescence assay can be implemented directly in the clarified lysate avoiding arduous purification and thereby increasing the throughput of designs or conditions that can be screened.

8. Expressing the protein cage with purification tags, while easing purification, could result in monomers that are assembly-inhibited on account of steric

blocking by the affinity label. Therefore, it is recommended to use purification strategies that involve either a very small tag (like His_6) or proteolytic removal of the tag. The small tag approach, although requiring controlled characterisation to ensure native assembly, may prove most flexible in that the same protein construct could be used for screening in homogeneously purified proteins, in lysates, and in living cells. This should be taken into account while designing the cloning strategy.

9. The ideal conditions for the expression of proteins varies depending on the nature of the protein. Care must be taken with protein cages that have been observed to exist as multiple, metastable oligomerization states. Extensive characterisation of the protein after expression should be performed to ensure that the desired oligomerization is achieved especially when establishing benchmarks in new systems.

10. Re-suspending the cell pellet in FlAsH buffer can save some steps and loss of protein especially if the screening will be undertaken with lysates. The addition of EDTA to the buffer can be helpful not only for maintaining ion concentration, but can also stop some protease activity in the lysate samples (see **Note 12**).

11. SEC provides information about what oligomerization states exist in each sample and their ratios with respect to each other. Thus, this technique can be used to ensure that the expression method has produced the expected oligomerization states of the protein cage. Extended purification methodologies, especially ones including digestion of purification tags can alter previously observed oligomerization state ratios, and re-characterisation is essential after purification. If the initial screens were performed in lysates, further

characterisation should be performed with purified proteins for the top hits. Moreover, SEC should prove to be a powerful characterisation tool in systems that form multiple oligomerization states of which the FlAsH binding sites were designed to probe one specifically. The FlAsH technique can be used in concert with SEC to determine which states are FlAsH-active. (29)

12. Purifying the protein into FlAsH buffer speeds up assaying as no buffer exchange is needed. Buffer exchange can cause protein aggregation or increase protease activity in a highly concentrated sample. Changing the concentration of the sample can also have an effect on the oligomerization state observed. Creating the simplest route from expression to screening is key to having the most reproducible results.

13. Storing lysate samples at 4 °C is recommended, however, temperature can possibly affect the oligomerization state ratio of the nanocage. Consistency between how all the samples are treated is important to maintain reproducibility.

14. Protein aggregation resulting in false positives can be the biggest problem at this step. The likelihood of this happening is increased if the design involves mutating residues in the more highly structured regions of the protein such as on a β-sheet.

15. Keep note of the sensitivity of the machine as different fluorimeters will provide different absolute readings; comparisons can be made by normalisation to controls. If filters are being used, ensure that they are the closest to the required wavelength and that their bandwidths do not overlap.

16. Protein concentration analysis for lysate samples is a measure of all the proteins in the sample. While it can be difficult to estimate how much of the sample is the protein of interest, steps such as PAGE gels and ensuring that the samples are treated identically from the protein production stage should lead to similar concentrations of the protein of interest. If multiple designs are to be compared, then additional control experiments should be conducted to ensure that each protein is expressed at a similar level.

17. In less stable nanocage systems, the oligomerization state may be altered if the sample is diluted from a stock concentration for the experiment. This also requires a control experiment.

18. The additives used in FlAsH fluorescence experiments can vary. TCEP is a commonly employed additive at concentrations ranging from 1 – 10 mM. It helps to maintain a reduced redox state. EDT is used to ensure that any FlAsH unbound to protein produces a low background. Commonly used concentrations range from 1 – 5 mM, however high concentrations have been shown to increase stringency and even inhibit FlAsH binding completely (*24*). 2-ME is used less often but it is suggested to not only help maintain the redox state but to increase the binding kinetics so that the FlAsH binding process can more readily reach equilibrium. Concentrations of 1 mM have typically been used. Optimisation for different systems is advised (*27*, *29*). (see Figure 3)

19. The amount of FlAsH dye to use can depend on the experiment and the sensitively of the fluorimeter. To quickly screen binding site designs with a yes/no output, a small amount of dye can be used (nM to µM range). For more quantitative experiments, the concentration of protein and therefore FlAsH

binding sites will have an effect on how much FlAsH-EDT_2 is added so as to maintain saturated stoichiometric ratios.

20. A standard incubation time between additions is 2 hours, however this can vary if the binding kinetics of the specific system are atypical.

21. Fluorescence experiments for purified proteins are the same as the above procedure, except that the amount of FlAsH-EDT_2 dye used might be altered to maintain saturated stoichiometry. Purified proteins provide much higher confidence that any observed fluorescence is coming from the designed binding site than in lysate experiments, but this can be countered by the decreased throughput due to the time required for purification.

22. The time required to fully denature the protein may also require optimisation. A positive control is especially useful here as often a small decrease in signal is observed in the presence of large amounts of Guanidine-HCl and this should be taken into account.

5. References

2. Kis, K., Volk, R. and Bacher, A. (1995) Biosynthesis of riboflavin - studies on the reaction-mechanism of 6,7-dimethyl-8-ribityllumazine synthase. *Biochemistry* 34, 2883-2892.
3. Walter, S. and Buchner, J. (2002) Molecular chaperones - Cellular machines for protein folding. *Angew Chem Intl Ed* 41, 1098-1113.
4. Homa, F. L. and Brown, J. C. (1997) Capsid assembly and DNA packaging in herpes simplex virus. *Rev Med Virol* 7, 107-122.
5. Aisen, P. and Listowsky, I. (1980) Iron transport and storage proteins. *Annu Rev Biochem* 49, 357-393.
6. Theil, E. C. (1987) Ferritin - structure, gene-regulation, and cellular function in animals, plants, and microorganisms. *Annu Rev Biochem* 56, 289-315.
7. Douglas, T., Strable, E., Willits, D., Aitouchen, A., Libera, M. and Young, M. (2002) Protein engineering of a viral cage for constrained nanomaterials synthesis. *Adv Mater* 14, 415-418.
8. Fan, R., Chew, S. W., Cheong, V. V. and Orner, B. P. (2010) Fabrication of Gold Nanoparticles Inside Unmodified Horse Spleen Apoferritin. *Small* 6, 1483-1487.

9. Fletcher, J. M., Harniman, R. L., Barnes, F. R. H., Boyle, A. L., Collins, A., Mantell, J., Sharp, T. H., Antognozzi, M., Booth, P. J., Linden, N., Miles, M. J., Sessions, R. B., Verkade, P. and Woolfson, D. N. (2013) Self-Assembling Cages from Coiled-Coil Peptide Modules. *Science* 340, 595-599.
10. Tinberg, C. E., Khare, S. D., Dou, J., Doyle, L., Nelson, J. W., Schena, A., Jankowski, W., Kalodimos, C. G., Johnsson, K., Stoddard, B. L. and Baker, D. (2013) Computational design of ligand-binding proteins with high affinity and selectivity. *Nature* 501, 212-216.
11. Fan, R., Boyle, A. L., Cheong, V. V., Ng, S. L. and Orner, B. P. (2009) A Helix Swapping Study of Two Protein Cages. *Biochemistry* 48, 5623-5630.
12. Ardejani, M. S., Li, N. X. and Orner, B. P. (2011) Stabilization of a Protein Nanocage through the Plugging of a Protein-Protein Interfacial Water Pocket. *Biochemistry* 50, 4029-4037.
13. Zhang, Y., Fu, J., Chee, S. Y., Ang, E. X. W. and Orner, B. P. (2011) Rational disruption of the oligomerization of the mini-ferritin E. coli DPS through protein-protein interface mutation. *Protein Sci* 20, 1907-1917.
14. Zhang, Y. and Orner, B. P. (2011) Self-Assembly in the Ferritin Nano-Cage Protein Superfamily. *Int J Mol Sci* 12, 5406-5421.
15. Ardejani, M. S., Chok, X. L., Foo, C. J. and Orner, B. P. (2013) Complete shift of ferritin oligomerization toward nanocage assembly via engineered protein-protein interactions. *Chem Commun* 49, 3528-3530.
16. Grueninger, D., Treiber, N., Ziegler, M. O. P., Koetter, J. W. A., Schulze, M.-S. and Schulz, G. E. (2008) Designed protein-protein association. *Science* 319, 206-209.
17. Ozawa, T. (2006) Designing split reporter proteins for analytical tools. *Anal Chim Acta* 556, 58-68.
18. Jares-Erijman, E. A. and Jovin, T. M. (2003) FRET imaging. *Nat Biotechnol* 21, 1387-1395.
19. Paulmurugan, R., Umezawa, Y. and Gambhir, S. S. (2002) Noninvasive imaging of protein–protein interactions in living subjects by using reporter protein complementation and reconstitution strategies. *Proc Natl Acad Sci* USA *USA* 99, 15608-15613.
20. Stains, C. I., Furman, J. L., Porter, J. R., Rajagopal, S., Li, Y., Wyatt, R. T. and Ghosh, I. (2010) A General Approach for Receptor and Antibody-Targeted Detection of Native Proteins Utilizing Split-Luciferase Reassembly. *ACS Chem Biol* 5, 943-952.
21. Miyawaki, A., Llopis, J., Heim, R., McCaffery, J. M., Adams, J. A., Ikura, M. and Tsien, R. Y. (1997) Fluorescent indicators for Ca2+ based on green fluorescent proteins and calmodulin. *Nature* 388, 882-887.
22. Gratzner, H. G. (1982) Monoclonal-antibody to 5-bromodeoxyuridine and 5-iododeoxyuridine - a new Reagent for Detection of DNA-replication. *Science* 218, 474-475.
23. Giepmans, B. N. G., Adams, S. R., Ellisman, M. H. and Tsien, R. Y. (2006) Review - The fluorescent toolbox for assessing protein location and function. *Science* 312, 217-224.
24. Neef, A. B. and Schultz, C. (2009) Selective Fluorescence Labeling of Lipids in Living Cells. *Angew Chem Intl Ed* 48, 1498-1500.
25. Griffin, B. A., Adams, S. R. and Tsien, R. Y. (1998) Specific covalent labeling of recombinant protein molecules inside live cells. *Science* 281, 269-272.

26. Stroffekova, K., Proenza, C. and Beam, K. G. (2001) The protein-labeling reagent FLASH-EDT2 binds not only to CCXXCC motifs but also non-specifically to endogenous cysteine-rich proteins. *Pflügers Arch* 442, 859-866.
27. Luedtke, N. W., Dexter, R. J., Fried, D. B. and Schepartz, A. (2007) Surveying polypeptide and protein domain conformation and association with FlAsH and ReAsH. *Nat Chem Biol* 3, 779-784.
28. Goodman, J. L., Fried, D. B. and Schepartz, A. (2009) Bipartite Tetracysteine Display Requires Site Flexibility for ReAsH Coordination. *Chembiochem* 10, 1644-1647.
29. Scheck, R. A., Lowder, M. A., Appelbaum, J. S. and Schepartz, A. (2012) Bipartite Tetracysteine Display Reveals Allosteric Control of Ligand-Specific EGFR Activation. *ACS Chem Biol* 7, 1367-1376.
30. Cornell, T. A., Fu, J., Newland, S. H. and Orner, B. P. (2013) Detection of Specific Protein–Protein Interactions in Nanocages by Engineering Bipartite FlAsH Binding Sites. *J Am Chem Soc* 135, 16618-16624.
31. Martin, B. R., Giepmans, B. N. G., Adams, S. R. and Tsien, R. Y. (2005) Mammalian cell-based optimization of the biarsenical-binding tetracysteine motif for improved fluorescence and affinity. *Nat Biotechnol* 23, 1308-1314.

Figure captions

Fig. 1. Conceptual evolution of FlAsH binding site design. (**A**) Structure and schematic of FlAsH-EDT_2. (**B**) Initial presentation of four cysteines on one face of an α-helix. (**C**) Optimized hairpin peptide. (**D**) Bipartite cysteine display with the two cysteine pairs placed proximately on each of the termini of the same protein. (**E**) Bipartite cysteine display with cysteine pairs straddling a protein-protein interface between interacting proteins. (**F**) Strategy to detect self-assembly of nanocage structures.

Fig. 2. Effect of the length and conformation of the sequence linking the nanocage protein Dps C-terminus to a bipartite pair. (Fluorescence is normalised to a positive control, G = glycine, A = alanine and P = proline), (1 μM FlAsH-EDT_2, 1 mM EDT, 1 mM 2-ME, 3.5 mM TCEP, 0.1 mg/ml final protein concentration). Results shown are from 6 replicates of lysate samples with over expressed proteins of interest. Error bars are S. D.

Fig. 3. Relationship between additives and the observed fluorescence intensity on a *E.coli* Bacterioferritin (Bfr) FerritinCCPGCC positive control (1 μM FlAsH-EDT_2, 1 mM EDT, 1 mM 2-ME, 3.5 mM TCEP, 0.1 mg/ml final protein concentration). Results shown are a summary of 6 repeats using purified proteins. Error bars are S. D.

Figure 1

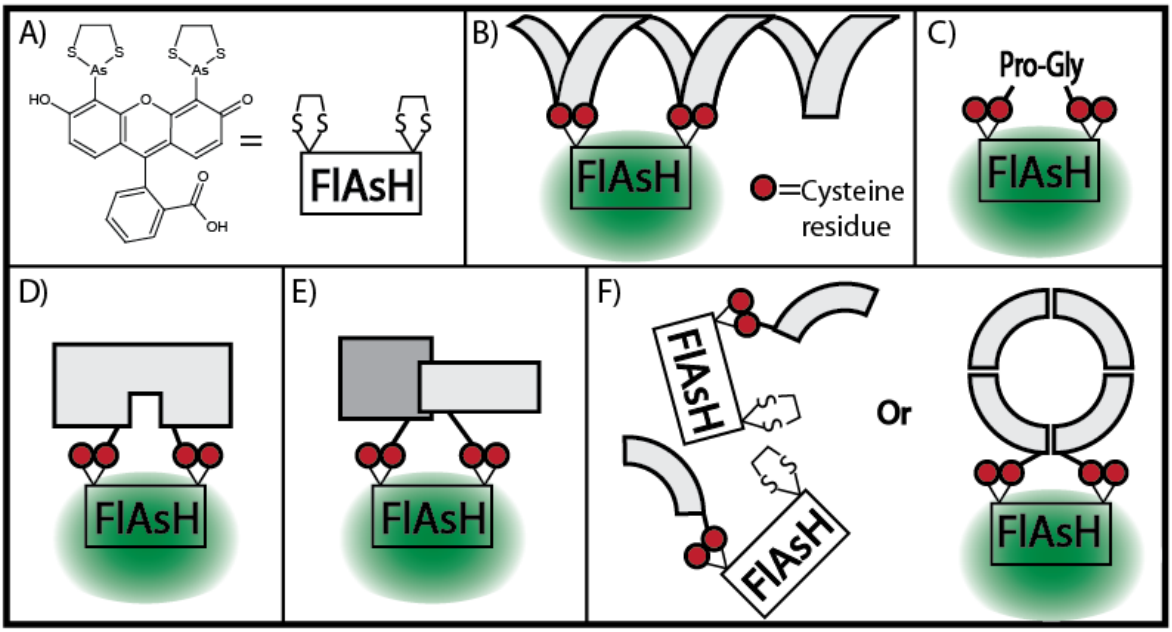


Figure 2

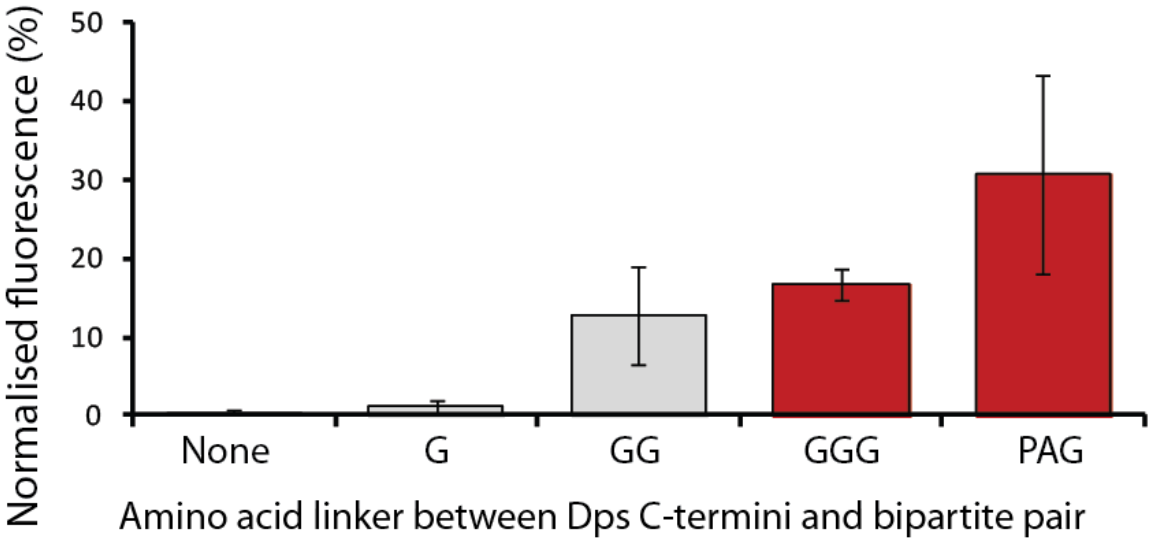
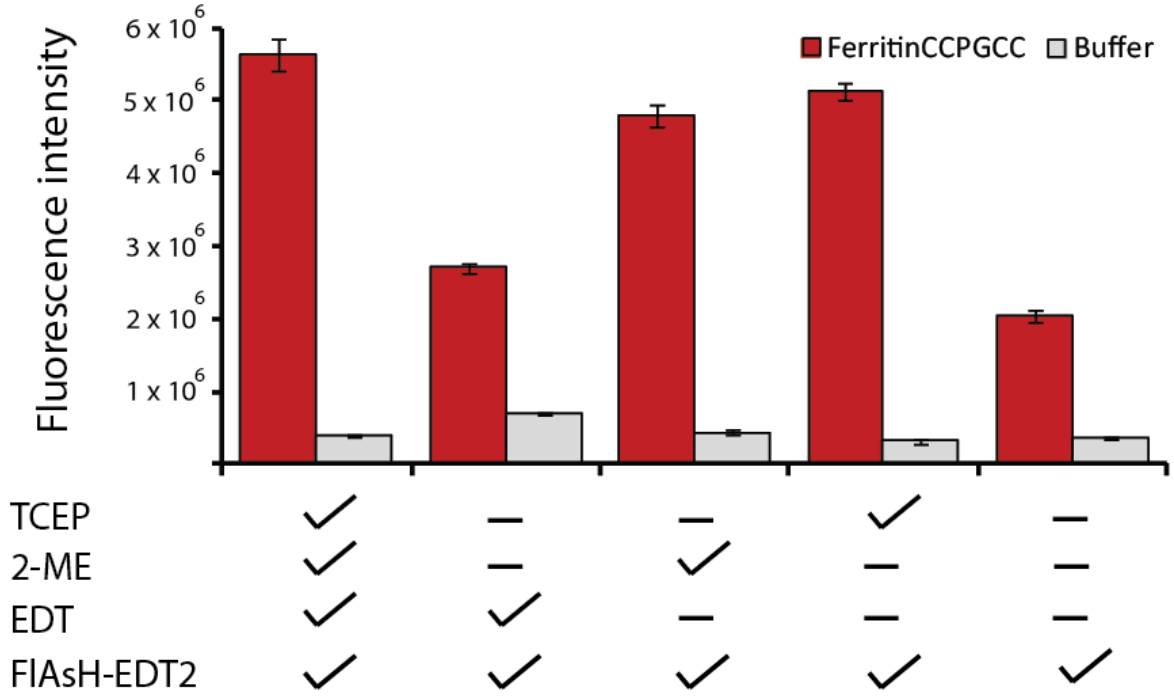


Figure 3


Fluorescence intensity
6 x 10^6
5 x 10^6
4 x 10^6
3 x 10^6
2 x 10^6
1 x 10^6
FerritinCCPGCC
Buffer
TCEP
2-ME
EDT
FlAsH-EDT2

Published work 2

Detection of Specific Protein-Protein Interactions in Nanocages by Engineering Bipartite FlAsH Binding Sites

Thomas A. Cornell[†, Δ], Jing Fu[Δ, ‡], Stephanie H. Newland[#, ‡], and Brendan P. Orner[†,*].

[†] Department of Chemistry, King's College London, London, SE1 1DB, United Kingdom. [Δ] Division of Chemistry and Biological Chemistry, Nanyang Technological University, Singapore, 637371. [#] School of Chemistry, University of Southampton, Southampton, SO17 1BJ, United Kingdom.

KEYWORDS Ferritin, DNA binding protein from starved cells, Protein cage, FlAsH, Bipartite tetracysteine display, Self-assembly, Protein engineering, Protein-protein interactions.

ABSTRACT: Proteins that form cage-like structures have been of much recent cross-disciplinary interest due to their application to bioconjugate and materials chemistry, their biological functions spanning multiple essential cellular processes, and their complex structure, often defined by highly symmetric protein-protein interactions. Thus, establishing the fundamentals of their formation, through detecting and quantifying important protein-protein interactions, could be crucial to understanding essential cellular machinery, and for further development of protein-based technologies. Herein we describe a method to monitor the assembly of protein cages by detecting specific, oligomerization state dependent, protein-protein interactions. Our strategy relies on engineering protein monomers to include cysteine pairs that are presented proximally if the cage state assembles. These assembled pairs of cysteines act as binding sites for the fluorescent reagent FlAsH which, once bound, provides a readout for successful oligomerization. As a proof of principle, we applied this technique to the iron storage protein, DNA-binding protein from starved cells (Dps) from *E.coli*. Several linker lengths and conformations for the presentation of the cysteine pairs were screened to optimize the engineered binding sites. We confirmed that our designs were successful in both lysates and with purified proteins, and that FlAsH binding was dependent upon cage assembly. Following successful characterization of the assay, its throughput was expanded. A two-dimension matrix of pH and denaturing buffer conditions was screened to optimize nanocage stability. We intend to use this method for the high throughput screening of protein cage libraries and of conditions for the generation of inorganic nanoparticles within the cavity of these and other cage proteins.

Introduction

Large, hollow, and often symmetric, cage-like protein assemblies, like ferritin nanocages and virus capsids, provide impetus for investigations into protein folding, protein-protein interactions and self-assembly, all of which underpin protein quaternary structure.[1,2,3] On a biofunctional level, similar protein nanostructures are involved in sequestering metals,[4] creating size-specific pockets of a hydrophobic environment to assist in protein folding,[3] catalyzing the generation of metabolites,[5] as well as delivering and protecting viral genomes.[6] Along with having fundamental importance, protein cages have been the focus of much applied research. To date, protein cages have been used as size constraining reaction vessels for the construction of inorganic materials, and for several potential bio-medically relevant applications such as drug and siRNA delivery.[7, 8, 9, 10, 11, 12, 13]

Many high resolution protein cage structures are available. [14, 15] These structures have paved the way for rational engineering and design, a pursuit that is important for enhancing the properties of protein cages to match those required for further applications. One successful design strategy, applied in different ways, has been to enhance the protein-protein interactions between monomers. [16, 17, 18, 19] However, this type of research can be protracted due to the necessity of iteratively purifying each mutant followed by extensive biophysical characterization with techniques that are often not directly related to cage formation.[2, 20]

Ferritins are ubiquitous protein cages whose structure has been extensively studied due to its relatively straightforward folding and assembly. They store cellular iron through mineralization inside their hollow cavity. *E.coli* DNA-binding protein from starved cells (Dps), a mini-ferritin, has a 9 nm outer diameter and assembles from twelve identical monomers.[21, 22, 23, 24, 25, 26] Each monomer folds into a four helix bundle with an additional helix along the loop between the second and third helix of the bundle (the "BC helix").[17, 27] During ferritin self-assembly, monomers rarely persist and for most ferritins, a 2-fold symmetric dimer is believed to be the most prevalent intermediate. This dimer is thought to be the fundamental building block for cage formation.[21] In Dps, these intermediates are most likely anti-parallel dimers. A consequence of this is that the termini of each monomer projects away from those of the other monomer; only with increased oligomerization to the fully formed cage state do the termini converge. (Figure 2A) Although ferritins have been pursued extensively for applications in material science,[10] only minimal work has been performed to optimize the properties of the ferritins for these applications.

Unfortunately, there are few methods for determining the oligomerization states of the cages other than expressing, purifying and assaying each protein followed by mostly low throughput biophysical techniques, such as size exclusion chromatography (SEC), dynamic light scattering (DLS) and transmission electron microscopy (TEM).[2, 16-17, 20, 28] Creating a system that can rapidly identify specific oligomerization states *in vivo* or in cell lysates, would greatly advance research on protein nanocages and protein self-assembly in general.

An ideal method would be one that could distinguish specific protein-protein interactions during the assembly process. A direct oligomerization assay employing a biarsinical fluorescent reagent,[29, 30] of which FlAsH and ReAsH are the most common, is one possibility. These reagents, which have been used as an alternative to GFP variants for protein labeling, [31, 32, 33, 34] exploit the affinity of their arsenic atoms for sulfur atoms in a protein. The reagents bind selectively to proteins with four appropriately presented cysteines (SI Figure S1). This binding results in fluorescence most likely due to a change in rotational properties about the carbon-arsenic bond.[35, 36] Originally, it was suggested that FlAsH ideally interacts with cysteines displayed on a single face of an alpha helix, but later it was shown that the sequence CCPGCC provides an ideal FlAsH binding site.

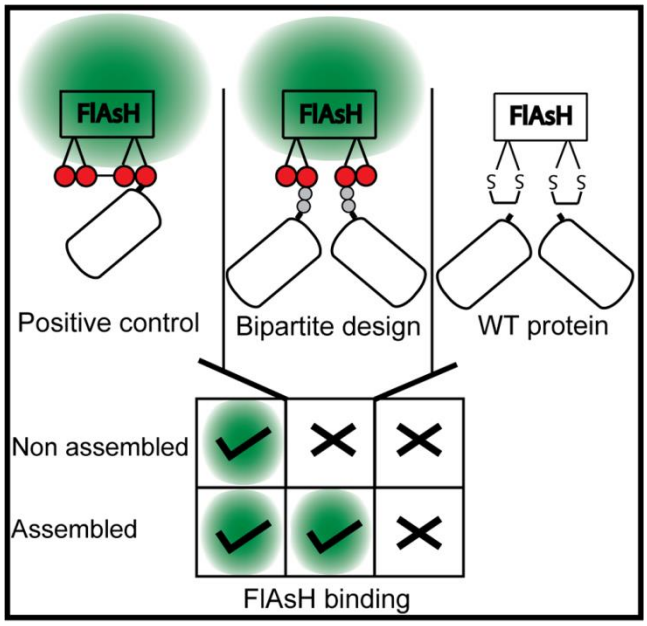


Figure 1 – Fluorescence (green) from the reagent, FlAsH, is assembly state dependent only in the bipartite design because of the need to have two sets of cysteine pairs (red) to form a binding site. The positive control provides a binding site that is neither folding, nor oligomerization dependent whereas the negative control provides no FlAsH binding. This is the basis of our strategy to directly monitor protein cage formation.

Recently, it has been shown that the four cysteines in the FlAsH binding motif can be split into two 'bipartite' cysteine pairs. If the two pairs are positioned apart from each other in the polypeptide sequence, FlAsH can be used to detect when they become proximal during protein folding. Similarly, the two pairs can be placed on two separate polypeptides, and FlAsH can be used to monitor the formation of a protein-protein interaction if the cysteines are positioned appropriately across the interface.[35, 37, 38] Recently this strategy was used to elucidate conformationally transduced signals through the cellular membrane and to provide an explanation for the divergent signaling outcomes of an EGF receptor that dimerizes through coiled-coil motifs.[39]

The first step in developing the biarsinical reagents into probes for protein cage assembly is to design binding sites that only appear upon cage formation. This goal is made more challenging by the fact that the monomers of protein cages, such as Dps, use multiple interfaces for cage formation and various states could be intermediates along the oligomerization pathway. Thus, a design with ultimate utility would be robust enough to distinguish between these oligomerization states. One possible solution therefore would be to design ferritin binding sites that exploit the divergence of monomer termini in the antiparallel dimer intermediate (see above). In the cage form, these termini converge. Thus, terminally appended cysteines would provide a sulfur rich area for a FlAsH binding site that only forms upon cage assembly and not in a dimer intermediate. One must also consider the location of native cysteine residues. Dps has a single cysteine close to the C-terminus of the monomer (SI Figure S2). While this will not impede this investigation, consideration must be taken when utilizing this approach for other proteins.

In our design, we placed cysteine pairs at the Dps C-terminus. (Figure 1) A series of mutants were generated to optimize the binding site; these mutants differ in how the two cysteines are displayed by the monomer either based on linker length (**DpsCC**, **DpsGCC**, **DpsGGCC**, **DpsGGGCC**) or conformation (**DpsPAGCC**). (Figure 2a) Extending the C-terminus to display the peptide sequence CCPGCC which includes all four cysteines of an ideal binding site on a single monomer (see above) resulted in the positive control (**DpsCCPGCC**) which would require neither folding nor assembly to generate a FlAsH signal. We used wild type Dps, which has no additional cysteines, as a negative control (**Dps**).

Results and Discussion

An oligomerization state assay with the most utility would be able to evaluate multiple samples rapidly. Therefore, instead of initially working with purified proteins, we used cellular lysates. This decision increased our throughput, allowing the screening of more potential designs. In addition, the uses of complex solutions allowed us to optimize for selectivity. In these conditions, (Figure 2b) the positive control, **DpsCCPGCC**, generated robust fluorescence with added FlAsH, and the negative control (**Dps**) gave nearly undetectable signal. Of the bipartite designs, **DpsCC** and **DpsGCC** had a very weak response whereas **DpsGGCC**, **DpsGGGCC**, and **DpsPAGCC** gave fluorescence that was between 15 and 35% of the positive control suggesting that a longer linker is ideal and one that is more rigid may be optimal. Observing signal in lysate samples shows this methods selectivity in the

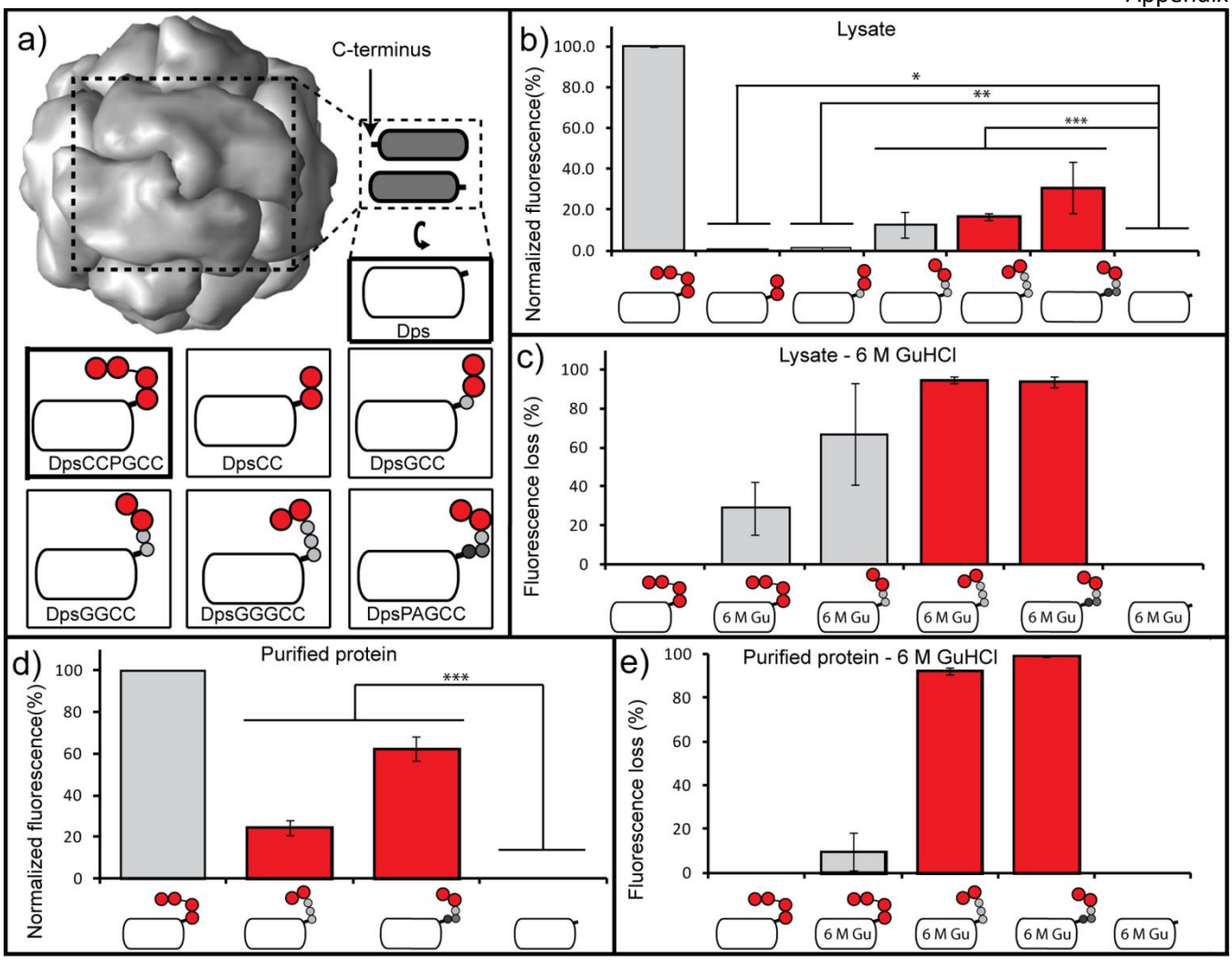


Figure 2 – (a) Dps crystal structure (PDB: 1DPS) emphasizing a dimer subunit and divergent presentation of the C-termini along with schematics representing proteins used in this study. The negative control (Dps) has no appended C-terminal cysteines whereas the monomer of the positive control (DpsCCPGCC) displays a full binding site. The other proteins present a pair of cysteines with a variety of flexible and constrained linkers. Cysteine (red), glycine (light grey), alanine (dark grey), proline (black) (b) FlAsH fluorescence, normalized to controls, for the proteins overexpressed as cell lysates diluted to the same total protein concentration (0.1 mg/ml). (c) Percent fluorescence loss for the proteins in lysates in denaturing conditions. (d) FlAsH fluorescence, normalized to controls, for the proteins overexpressed and purified, diluted to the same protein concentration (0.1 mg/ml) (e) Percent fluorescence loss for the purified proteins in denaturing conditions. The lysate and purified protein data are from six and eight replicates respectively. Error bars are S.D. * Two-tailed P-values = 0.4804. ** Two-tailed P-values = 0.020050. *** Two-tailed P-values = 0.0001.

presence of many ions, organic molecules and proteins. While lysates are still not as complex as intact cells, this report shows positive movement towards that direction. This investigation uses set concentrations of 1, 2 ethanedithiol (EDT), 2-mercaptoethanol (2-ME), tris(2-carboxyethyl)phosphine (TCEP) and ethylenediaminetetraacetic acid (EDTA) and while TCEP has been shown to be necessary to ensure that there are no disulfide bonds, the methodology for EDT and 2-ME is less clear with the latter only been used sparingly. [29] The EDT ligand is needed to keep unbound FlAsH non-fluorescent, however its concentration is often modified in line with the system being used.[30, 37, 38] Moreover EDTA which is commonly used could be doubly benefiting to lysate experiments by removing the extra ions seen. All of these additives would require further optimization if a high through put screen was to be attempted.

To confirm that the FlAsH binding is indeed dependent on cage formation, we subjected the lysates to similar experiments but in denaturing conditions. (Figure 2c) As expected the negative control, **Dps**, exhibited no change in fluorescence upon denaturation. The positive control, **DpsCCPGCC**, was also expected to have no change, however, it did generate a small, but significant loss which may be attributed to restricted access to the tetracysteine tag in the cage and may be suggestive that more optimized

equilibration times could enhance the assay. As expected, the

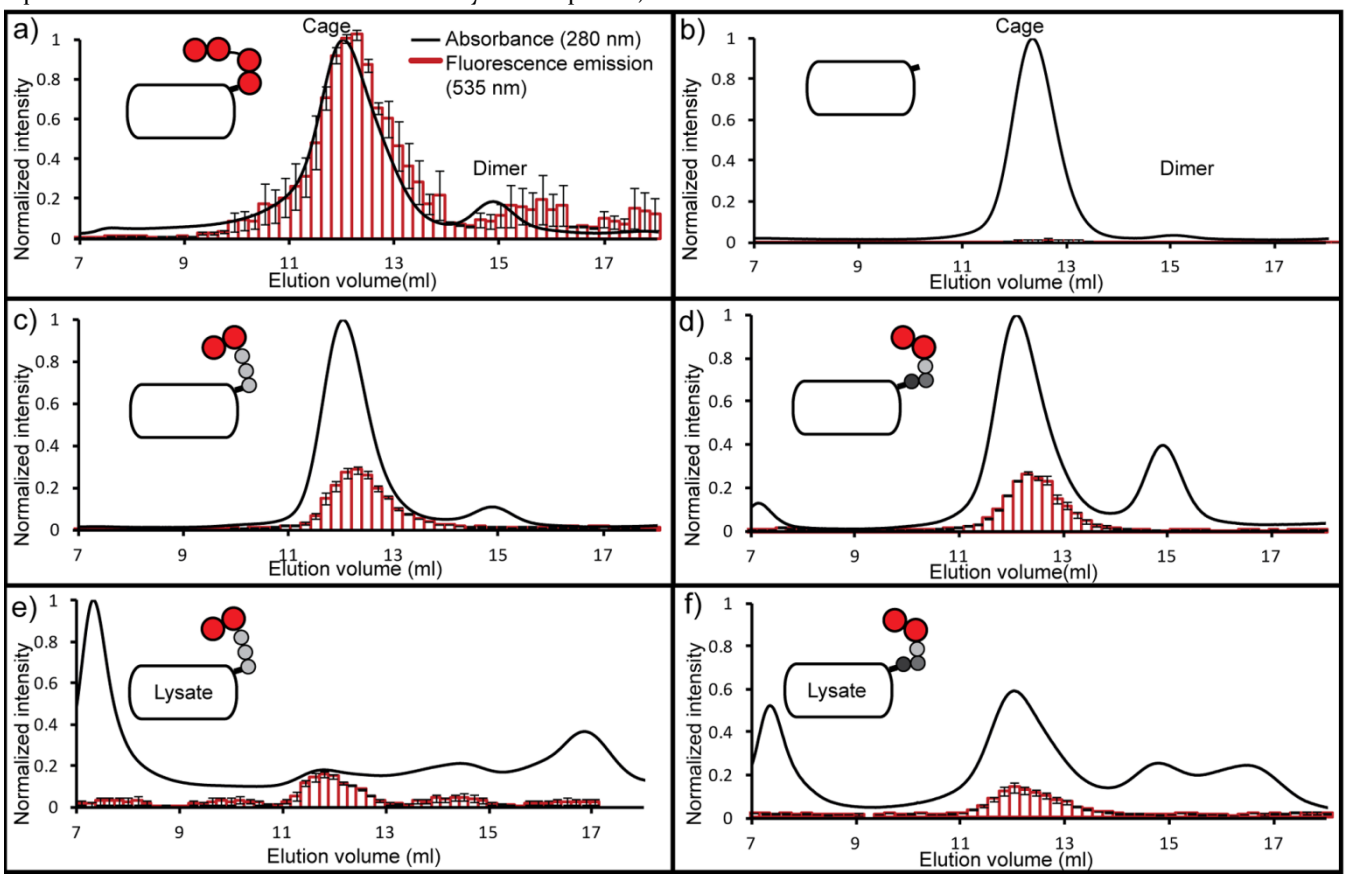


Figure 3 – Size exclusion chromatograms monitored at absorbance of 280 nm (protein) and fluorescence at 535 nm (bound FlAsH) to determine if the assay is selective for the cage oligomerization state. (a-d) Purified proteins (0.5 ml injection of 1 mg/ml). (e, f) Lysate solutions with overexpressed designed protein (0.5 ml injection of 1 mg/ml total protein). All absorbance and fluorescence signals are normalized to the positive control. Results for both the UV and fluorescence traces are averages of three runs.

bipartite designs displayed a large loss in fluorescence upon denaturation with **DpsGGGCC** and **DpsPAGCC** exhibiting nearly a 100% loss in signal with added denaturant. To verify these results, the leading proteins were expressed, purified (SI Figure S11 for sequencing data) and fully characterized (SI Figure S4-8 and SI Table S4 for gels and mass spec, SEC, TEM, CD) and were subjected to the binding experiment and subsequent denaturation. (Figures 2d and 2e) Again, the controls behaved as expected with the exception of the positive control that, again, lost a small amount of signal upon denaturation.

The bipartite designs, **DpsGGGCC** and **DpsPAGCC**, displayed strong, oligomerization dependent binding of FlAsH with the latter generating ~50% of control. While the stoichiometry between dye and protein is unclear in the lysate samples, as our set protein concentrations are a measure of all soluble proteins, it is clear in the purified experiments that the sum of proteins and therefore binding sites, far outnumber the amount of dye. While this could be of concern in a quantitative screen, this does not detract away from this investigation to discover binding sites that work across complex ferritin interfaces.

The steady state fluorescence experiments, coupled with denaturation, strongly demonstrate that we have successfully designed FlAsH binding sites that are oligomerization dependent. While those experiments, combined with the geometrical placement of the binding sites, suggested that this dependence is specific to the cage state over the dimer state, it needed to be confirmed directly. We have previously shown that some ferritins can exist in solution as mixtures of cage and dimer through the use of SEC.[28] Therefore we employed SEC, combining protein absorbance (280 nm) with FlAsH fluorescence (535 nm). If the binding sites were successfully designed to appear during the formation of specific protein-protein interactions, fluorescent peaks should correspond to only certain oligomerization forms. (Figure 3) The negative control, **Dps**, which is a wild type protein we have worked with extensively, generates a single peak from the cage and no detectable dimer when monitored at 280 nm. (Figure 3b) Also as expected, no peak is observed in the FlAsH channel. For the positive control, **DpsCCPGCC**, a cage and a small dimer peak at 280 nm are observed, and the fluorescence elutes with both, again as expected, indicating no preference for oligomerization state. (Figure 3a) The proteins with the bipartite binding sites,

DpsPAGCC and **DpsGGGCC**, both show a cage and a dimer peak at 280 nm however, the fluorescence only elutes with the cage, suggesting that the designed binding sites are forming only upon cage formation and that this assay can distinguish between oligomerization states. (Figures 3c and 3d) It should be noted that **DpsPAGCC** did aggregate slightly. The fact that the aggregate peak also had no fluorescence further demonstrates the robustness of this approach. As a further test of specificity, the same technique was used to assess clarified lysates for **DpsPAGCC** and **DpsGGGCC**. (Figures 3e and 3f) Again, fluorescence elutes at a volume consistent with the cage state. Taken together these data strongly suggest that not only are the designed binding sites forming upon oligomerization, they form only upon cage formation.

An additional goal of this technique is not only to develop a direct screen for specific oligomerization states of protein cages, but also to expand the throughput of this characterization. Therefore it was modified to a 96-well plate format. As a proof of principle, but also as a means to optimize our work with these proteins (see above), we sought to discover conditions of highest stability. First we expanded our denaturation experiment described in Figure 1 to screen a variety of denaturant concentrations. (Figure 4a) As expected, and consistent with our previous data, the positive control, **DpsCCPGCC**, generated high signal at all conditions, demonstrating that FlAsH binding is not oligomerization dependent. The negative control, **Dps**, generates no signal across all conditions. However, a protein with the designed bipartite binding site, **DpsPAGCC**, loses its ability to bind FlAsH near 1.4M denaturant. (SI Figure S10)

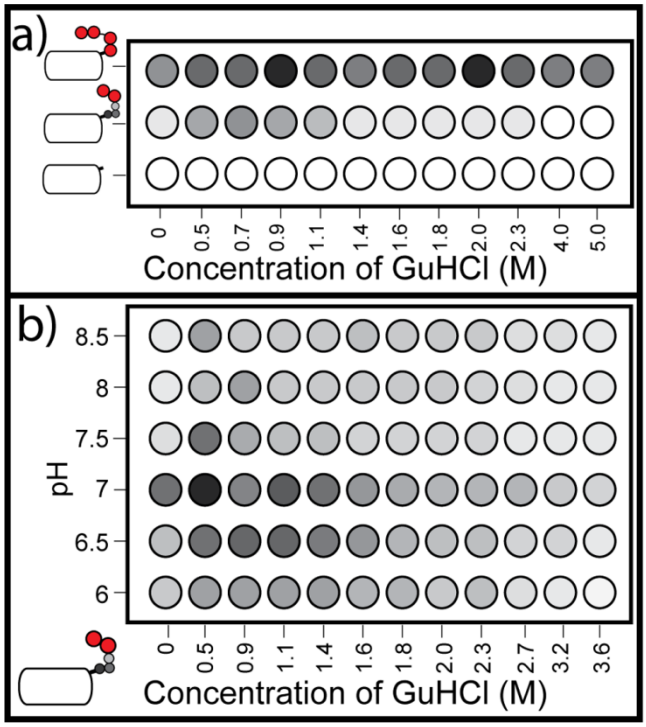


Figure 4 – Medium throughput screen of buffer conditions favoring cage oligomerization state. Fluorescence intensity is greyscaled so that black is most intense and white is least. (a) Normalized fluorescence intensity of **DpsCCPGCC**, **DpsPAGCC** and **Dps** as a function of GuHCl concentration at pH 7.8. Values are the average of at least six replicates. S.D. is included in Supporting Information (SI Figure S10) (b) Normalized fluorescence intensity for **DpsPAGCC** as a function of both pH and GuHCl. Values are averages of three replicates. S.D. is included in Supporting Information (SI Figure S9)

Further expanding our enquiry into ideal conditions for Dps stability, we screened both denaturant and pH concurrently. (Figure 4b and SI Figure S9) It should be noted that, consistent with most of the literature,[35,37] we usually preform our experiments with Dps at pH 7.8. Thus, it was somewhat surprising to discover that the protein is most stable below pH 7.0. These experiments would have been less possible in lower throughput screens as they would have taken longer and expended a large amount of reagents. This not only emphasizes the strength of this approach in that its throughput can be expanded, but this increased throughput can lead to useful experiments (with potentially unexpected results) that may not have been run due to the exigent nature of traditional techniques.

Conclusion

Protein cages have the potential for applications in fields as wide spanning as drug delivery, catalysis, and nanomaterials. Moreover, they can act as model systems to study biologically ubiquitous protein-protein interactions, self-assembly and quaternary structure all of which are at the cutting edge of pharmaceutical interest. However, for protein cages to reach their full potential, tools to assess directly their solution assembly properties in higher throughput are necessary. By engineering FlAsH binding sites and protein-protein interfaces that only form in the cage oligomerization state and not in the presumed dimer intermediate, we designed a system to directly detect assembly in cellular lysates of the miniferritin *E.coli* Dps. We have shown that indeed this system is oligomerization dependent and is specific for the cage and not the dimer. In addition, we demonstrated the scalability of this system by performing a two dimensional, medium throughput screen to determine conditions that favor the cage state. We are currently expanding this technique to other ferritins in the form of protein libraries screened in whole bacteria for the purpose of discovering "switchable" protein cages and those with bespoke properties for specific applications like nanoparticle formation while optimizing the designs based on ideal binding sites. [38] In addition, we are intending to use this technique to target specific oligomerization states in order to monitor their formation along the self-assembly pathway. We believe that this strategy can be easily ported to other protein cages and self-assembling protein systems in order to optimize their properties and to understand their formation.

Methods and Materials

FlAsH Binding in Lysates. The pET-22b expression vectors containing Dps design variants (SI Methods and materials S1 for cloning information) were transformed into Rosetta *E.coli* cells (Novagen) and plated on LB agar plates (50 µl/ml of carbenicillin and 34 µl/ml chloramphenicol). Selected colonies then were grown in LB (3 ml with 50 µl/ml of

carbenicillin, 37 °C, overnight) as a pre-culture which was then added to LB (100 ml) and grown (37 °C) until an $O.D_{600}$ of 0.6. Protein expression was then induced by the addition of IPTG (50 µl of a 1 M stock) and the culture was further incubated (3 h, 30 °C). The cells were isolated by centrifugation (4,000 rpm, 15 min at 4 °C). The cell pellet was resuspended in FlAsH buffer (100 mM Tris.HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) and sonicated (Misonix, Ultrasonic cell disruptor, pulsed 5 s on 5 s off for 5 min). The protein solution was clarified by centrifugation (15,000 rpm, 45 min, 4 °C) and then filtered (Sartorius, 0.2 µm).

The protein concentration was determined (BCA, Novagen) and the cell lysate was diluted to 1 mg/ml with FlAsH buffer. To insure that all the 1 mg/ml samples had similar amounts of the desired protein, each was analyzed by SDS PAGE (SI Figure S3). Each fluorescence experiment contained protein lysate (200 µl, 0.1 mg/ml) in FlAsH buffer, TCEP (Sigma, final concentration of 3.5 mM), EDT (Sigma, final concentration of 1 mM) and 2-ME (Sigma, final concentration of 1 mM) were added and the solution was incubated (room temperature, 2 h) followed by the addition of FlAsH-EDT_2 (Invitrogen, final concentration of 0.1 µM) followed by a further incubation (room temperature, 2 h) in the dark. Each lysate sample was tested in a black Corning 96 well plate in a PerkinElmer Envision 2101 multilabel plate reader, with each design being expressed three times and each expression being tested in 6 different wells and re read twice (Ex filter 485 nm bandwidth 14 nm, Em filter 535 nm bandwidth 25 nm). For the denatured experiments, the above was repeated but with the extra addition of 6 M guanidine and incubated for 2 hours prior to the addition of TCEP, EDT and 2-ME, with protein concentration remaining the same as in previous un-denatured experiments.

FlAsH Binding with Purified Proteins. The pET-32b vectors containing the Dps variants (SI Methods and materials S2 for cloning information) were transformed into Rosetta *E.coli* cells (Novagen) and plated on LB agar plates (50 µl/ml of carbenicillin and 34 µl/ml of chloramphenicol). Selected colonies were then grown in LB (5 ml, 37 °C, overnight) as a pre-culture which was added to LB (500 ml) and grown (37 °C) until an $O.D_{600}$ of 0.6. Protein expression was then induced by the addition of IPTG (250 µl of a 1 M stock) and the cultures were further incubated (3 h, 30 °C). The cells were isolated by centrifugation (4,000 rpm, 20 min, 4 °C). The cell pellet was resuspended in lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 40 mM Imidazole, 1 mM EDTA, pH 8). Cellytic (10x, Sigma) was added and the solution was incubated (20 min, on ice) and then sonicated (Misonix, ultrasonic cell distruptor, pulsed 5 s on, 5 s off for 5 min). The protein solution was clarified by centrifugation (15,000 rpm, 45 min at 4 °C) and then filtered (Sartorius, 0.2 µm). The protein was purified via affinity purification (GE, Histrap FF, 5 ml, (wash buffer-40 mM Imidazole, 50 mM NaH_2PO_4, 300 mM NaCl, pH 7.4), (elution buffer-500 mM Imidazole, 50 mM NaH_2PO_4, 300 mM NaCl, pH 7.4)). Enterokinase digestion (NEB 2 µg/ml) was performed to cleave off the peptide tag from the protein of interest followed by a second Histrap (GE, Histrap HP, 5 ml, (wash buffer-50 mM NaH_2PO_4, 300 mM NaCl, pH 7.4), (elution buffer-500 mM Imidazole, 50 mM NaH_2PO_4, 300 mM NaCl, pH 7.4)) to remove the tag from solution. The protein solution was further purified by size exclusion chromatography (GE Hiload 16/60 Superdex, running buffer-50 mM NaH_2PO_4). Each protein was then placed into FlAsH buffer via ultrafiltration (Millipore).

Each purified protein was tested by incubating the protein (200 µl of 0.1 mg/ml) in FlAsH buffer with TCEP (Sigma, final concentration of 3.5 mM), EDT (Sigma, final concentration of 1 mM) and 2-ME (Sigma, final concentration of 1 mM) and left to incubate (2 h, room temperature) followed by the addition of FlAsH-EDT_2 (Invitrogen, final concentration of 0.1 µM, 2 h, room temperature). Each sample was prepared directly into a black Corning 96 well plate which was tested in a PerkinElmer Envision 2101 multilabel plate reader (Ex filter 485 nm bandwidth 14 nm, Em filter 535 nm bandwidth 25 nm). Each pure protein was tested six times and reread three times. For the denaturant experiment, each protein was tested as described above but with the addition of 6 M guanidine, while keeping the protein concentration the same and was incubated for 2 hours prior to the addition of TCEP, 2-ME, and EDT.

Analytical Size Exclusion Chromatography (SEC). The samples (0.5 ml of 1 mg/ml) in FlAsH buffer (100 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, pH 7.8) were injected onto the column (GE Superdex 200 10/300 GL) at 0.5 ml/min with each protein repeated 3 times. The column was calibrated using six proteins as standards (GE Biosystems Calibration Kit). (See SI Figure S7.)

For fluorescent monitoring, samples were prepared by adding to pure protein (0.5 ml of 1 mg/ml) in FlAsH buffer, TCEP (Sigma, final concentration of 3.5 mM), EDT (Sigma, final concentration of 1 mM) and 2-ME (Sigma, final concentration of 1 mM) and incubated (2 h, room temperature). FlAsH-EDT_2 dye was added and incubated in the dark (Invitrogen, final concentration of 0.4 µM, 2 h, room temperature). This sample was desalted (GE, HiTrap 5 ml Desalting column) before injection onto a SEC column (GE Superdex 200 10/300 GL). Samples were taken every 200 µl during the elution and placed into a black Corning 96 well plate. This plate was tested in a PerkinElmer Envision 2101 multilabel plate reader (Ex filter 485 nm bandwidth 14 nm, Em filter 535 nm bandwidth 25 nm). Each purified protein was tested three times with each plate reread three times.

Transmission Electron Microscopy (TEM). TEM was performed on a FEI, Tecnai G^2 20, electron microscope set at 200 KeV. Proteins were immobilized on Formvar/carbon coated 3.05 mm copper grids (TAAB) and negatively stained with 1 % Uranyl acetate. [7] Micrographs were analyzed using ImageJ. [41] (See SI FigureS8.)

Circular Dichroism Spectroscopy (CD). The purified proteins were used in FlAsH buffer (100 mM Tris, 100 mM NaCl, 1 mM EDTA, pH 7.8). This experiment was performed on a Applied Photophysics LTD Chirascan spectrometer in a range of 200 nm to 260 nm with a protein concentration of 0.2 mg/ml with a path length of 0.5 mm. Thermal melts were performed on all purified proteins (0.2 mg/ml in FlAsH buffer) in a range of 4 to 85 degrees Celsius. [42, 43] (See SI Figure S4, S5 and Table S4.)

Electrospray Mass Spectrometry. Purified proteins were desalted (GE, Hitrap 5ml Desalting column) and analyzed on a Bruker MaXis mass spectrometer after the addition of 1 % Formic acid.(See SI Table S5.)

pH Vs GuHCl measurement. Each purified protein was analyzed by incubating the protein (200 µl of 0.1 mg/ml) in a buffer at the relevant pH (pH 6-7, 100 mM Citrate-phopshate. pH 7.5-8.5 100 mM Tris.HCl) and with the correct concentration of GuHCl, with TCEP (Sigma, final concentration of 3.5 mM), EDT (Sigma, final concentration of 1 mM) and 2-ME (Sigma, final concentration of 1 mM) and incubated (2 h, room temperature) followed by the addition of FlAsH-EDT_2 (Invitrogen, final concentration of 0. 1 µM, 2 h, room temperature). Each sample was prepared directly in a black Corning 96 well plate which was scanned in a PerkinElmer Envision 2101 multilabel plate reader (Ex filter 485 nm bandwidth 14 nm, Em filter 535 nm bandwidth 25 nm). Each protein was prepared three times separately and each was reread three times.[44]

ASSOCIATED CONTENT

Supporting Information. This includes SDS-PAGE for lysate samples, SDS-PAGE for all purified proteins as well as TEM, CD, SEC, and electrospray mass spectrometry data for their characterization and sequence alignments and primers used in their manufacture. Raw data for the pH vs GuHCl experiment is also found here. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*brendan.orner@kcl.ac.uk.

Author Contributions

‡These authors contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

We thank E. Hobart, N. Luedtke, Y. Zhang, M. Ardejani, and M. Peczuh for insightful conversations, Fan Rongli for the plasmid coding for WT Dps in pET-32b, and A. T. Phan for the encouragement to perform the experiment in Figure 3. We also thank KCL's Centre for Ultrastructural Imagining and Centre for Bimolecular Spectroscopy for help in characterization of the proteins, D. Thurston's lab for help and access to instrumentation, and J. McDonnell for help with the ESI. T.A.C was sponsored by SINGA and BSE scholarships at NTU and King's respectively and J.F. was supported by CBC's undergraduate research fund. The research at NTU was supported by an SPMS start-up grant, a Singapore Ministry of Education Academic Research Fund Tier 1 Grant (RG 53/06) and B.P.O's personal salary. At KCL it was supported by start up funds from the School of Biomedical Sciences.

ABBREVIATIONS

Dps, DNA-binding protein from starved cells; WT Dps, wild type DNA-binding protein from starved cells; FlAsH-EDT_2, 4',5'-bis(1,3,2-dithioarsolan-2-yl)fluorescein-(1,2-ethanedithiol)2; GU·HCl, Guanidine hydrochloride; Tris-HCl, tris(hydroxymethyl)aminomethane hydrochloride; EDT, 1,2 ethanedithiol; TCEP, tris(2-carboxyethyl)phosphine; 2-ME, 2-mercaptoethanol; SEC, size exclusion chromatography; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis; GFP, green fluorescent protein; PCR, polymerase chain reaction; LB, Luria broth; EDTA, Ethylenediaminetetraacetic acid; BCA, bicinchoninic acid; IPTG, Isopropyl β-D-1-thiogalactopyranoside; Ek, Enterokinase; TEM, transmission electron microscopy; CD, circular dichroism; FPR, Fluorescent plate reader; GE, General Electric; Ex, Excitation; Em, Emission; NEB, New England Bioscience.

REFERENCES

1. Grueninger, D.; Treiber, N.; Ziegler, M. O. P.; Koetter, J. W. A.; Schulze, M.-S.; Schulz, G. E., Designed protein-protein association. Science 2008, 319 (5860), 206-209.
2. Fan, R.; Boyle, A. L.; Cheong, V. V.; Ng, S. L.; Orner, B. P., A Helix Swapping Study of Two Protein Cages. Biochemistry 2009, 48 (24), 5623-5630.
3. Walter, S.; Buchner, J., Molecular chaperones - Cellular machines for protein folding. Angewandte Chemie-International Edition 2002, 41 (7), 1098-1113.
4. Aisen, P.; Listowsky, I., Iron transport and storage proteins. Annual Review of Biochemistry 1980, 49, 357-393.
5. Kis, K.; Volk, R.; Bacher, A., Biosynthesis of riboflavin - studies on the reaction-mechanism of 6,7-dimethyl-8-ribityllumazine synthase. Biochemistry 1995, 34 (9), 2883-2892.
6. Homa, F. L.; Brown, J. C., Capsid assembly and DNA packaging in herpes simplex virus. Reviews in Medical Virology 1997, 7 (2), 107-122.
7. Fan, R.; Chew, S. W.; Cheong, V. V.; Orner, B. P., Fabrication of Gold Nanoparticles Inside Unmodified Horse Spleen Apoferritin. Small 2010, 6 (14), 1483-1487.
8. Zhang, L.; Swift, J.; Butts, C. A.; Yerubandi, V.; Dmochowski, I. J., Structure and activity of apoferritin-stabilized gold nanoparticles. Journal of Inorganic Biochemistry 2007, 101 (11-12), 1719-1729.
9. Ueno, T.; Suzuki, M.; Goto, T.; Matsumoto, T.; Nagayama, K.; Watanabe, Y., Size-selective olefin hydrogenation by a Pd nanocluster provided in an apo-ferritin cage. Angewandte Chemie-International Edition 2004, 43 (19), 2527-2530.
10. Kramer, R. M.; Li, C.; Carter, D. C.; Stone, M. O.; Naik, R. R., Engineered protein cages for nanomaterial synthesis. Journal of the American Chemical Society 2004, 126 (41), 13282-13286.
11. Woersdoerfer, B.; Woycechowsky, K. J.; Hilvert, D., Directed Evolution of a Protein Container. Science 2011, 331 (6017), 589-592.
12. Keyes, J. D.; Hilton, R. J.; Farrer, J.; Watt, R. K., Ferritin as a photocatalyst and scaffold for gold nanoparticle synthesis. Journal of Nanoparticle Research 2011, 13 (6), 2563-2575.

13. Qiu, H. J.; Dong, X. C.; Sana, B.; Peng, T.; Paramelle, D.; Chen, P.; Lim, S., Ferritin-Templated Synthesis and Self-Assembly of Pt Nanoparticles on a Monolithic Porous Graphene Network for Electrocatalysis in Fuel Cells. Acs Applied Materials & Interfaces 2013, 5 (3), 782-787.

14. Grant, R. A.; Filman, D. J.; Finkel, S. E.; Kolter, R.; Hogle, J. M., The crystal structure of Dps, a ferritin homolog that binds and protects DNA. Nature Structural Biology 1998, 5 (4), 294-303.

15. Luo, J.; Liu, D.; White, M. A.; Fox, R. O., 1F33: The structural basis for dna protection by e. coli dps protein. Protein Data Bank 2003.

16. Ardejani, M. S.; Li, N. X.; Orner, B. P., Stabilization of a Protein Nanocage through the Plugging of a Protein-Protein Interfacial Water Pocket. Biochemistry 2011, 50 (19), 4029-4037.

17. Ardejani, M. S.; Chok, X. L.; Foo, C. J.; Orner, B. P., Complete shift of ferritin oligomerization toward nanocage assembly via engineered protein-protein interactions. Chemical Communications 2013, 49 (34), 3528-3530.

18. Fletcher, J. M.; Harniman, R. L.; Barnes, F. R. H.; Boyle, A. L.; Collins, A.; Mantell, J.; Sharp, T. H.; Antognozzi, M.; Booth, P. J.; Linden, N.; Miles, M. J.; Sessions, R. B.; Verkade, P.; Woolfson, D. N., Self-Assembling Cages from Coiled-Coil Peptide Modules. Science 2013, 340 (6132), 595-599.

19. Woolfson, D. N., The design of coiled-coil structures and assemblies. Fibrous Proteins: Coiled-Coils, Collagen and Elastomers 2005, 70, 79-+.

20. Zhang, Y.; Raudah, S.; Teo, H.; Teo, G. W. S.; Fan, R.; Sun, X.; Orner, B. P., Alanine-shaving Mutagenesis to Determine Key Interfacial Residues Governing the Assembly of a Nano-cage Maxi-ferritin. Journal of Biological Chemistry 2010, 285 (16), 12078-12086.

21. Zhang, Y.; Fu, J.; Chee, S. Y.; Ang, E. X. W.; Orner, B. P., Rational disruption of the oligomerization of the mini-ferritin E. coli DPS through protein-protein interface mutation. Protein Science 2011, 20 (11), 1907-1917.

22. Haile, D. J., Regulation of genes of iron metabolism by the iron-response proteins. American Journal of the Medical Sciences 1999, 318 (4), 230-240.

23. Harrison, P. M.; Arosio, P., Ferritins: Molecular properties, iron storage function and cellular regulation. Biochimica Et Biophysica Acta-Bioenergetics 1996, 1275 (3), 161-203.

24. Hintze, K. J.; Theil, E. C., Cellular regulation and molecular interactions of the ferritins. Cellular and Molecular Life Sciences 2006, 63 (5), 591-600.

25. Leipuviene, R.; Theil, E. C., The family of iron responsive RNA structures regulated by changes in cellular iron and oxygen. Cellular and Molecular Life Sciences 2007, 64 (22), 2945-2955.

26. Theil, E. C., Ferritin - structure, gene-regulation, and cellular function in animals, plants, and microorganisms. Annual Review of Biochemistry 1987, 56, 289-315.

27. Fan, R.; Boyle, A. L.; Vee, V. C.; See, L. N.; Orner, B. P., A helix swapping study of two protein cages. Biochemistry 2009, 48 (24), 5623-5630.

28. Zhang, Y.; Orner, B. P., Self-Assembly in the Ferritin Nano-Cage Protein Superfamily. International Journal of Molecular Sciences 2011, 12 (8), 5406-5421.

29. Griffin, B. A.; Adams, S. R.; Tsien, R. Y., Specific covalent labeling of recombinant protein molecules inside live cells. Science 1998, 281 (5374), 269-272.

30. Griffin, B. A.; Adams, S. R.; Jones, J.; Tsien, R. Y., Fluorescent labeling of recombinant proteins in living cells with FlAsH. Applications of Chimeric Genes and Hybrid Proteins Pt B 2000, 327, 565-578.

31. Tsien, R. Y., The green fluorescent protein. Annual Review of Biochemistry 1998, 67, 509-544.

32. Adams, S. R.; Campbell, R. E.; Gross, L. A.; Martin, B. R.; Walkup, G. K.; Yao, Y.; Llopis, J.; Tsien, R. Y., New biarsenical Ligands and tetracysteine motifs for protein labeling in vitro and in vivo: Synthesis and biological applications. Journal of the American Chemical Society 2002, 124 (21), 6063-6076.

33. Ignatova, Z.; Gierasch, L. M., Monitoring protein stability and aggregation in vivo by real-time fluorescent labeling. Proceedings of the National Academy of Sciences of the United States of America 2004, 101 (2), 523-528.

34. Ray-Saha, S.; Schepartz, A., Visualizing Tyrosine Kinase Activity with Bipartite Tetracysteine Display. Chembiochem 2010, 11 (15), 2089-2091.

35. Scheck, R. A.; Schepartz, A., Surveying Protein Structure and Function Using Bis-Arsenical Small Molecules. Accounts of Chemical Research 2011, 44 (9), 654-665.

36. Madani, F.; Lind, J.; Damberg, P.; Adams, S. R.; Tsien, R. Y.; Graslund, A. O., Hairpin Structure of a Biarsenical-Tetracysteine Motif Determined by NMR Spectroscopy. Journal of the American Chemical Society 2009, 131 (13), 4613-+.

37. Luedtke, N. W.; Dexter, R. J.; Fried, D. B.; Schepartz, A., Surveying polypeptide and protein domain conformation and association with FlAsH and ReAsH. Nature Chemical Biology 2007, 3 (12), 779-784.

38. Goodman, J. L.; Fried, D. B.; Schepartz, A., Bipartite Tetracysteine Display Requires Site Flexibility for ReAsH Coordination. Chembiochem 2009, 10 (10), 1644-1647.

39. Scheck, R. A.; Lowder, M. A.; Appelbaum, J. S.; Schepartz, A., Bipartite Tetracysteine Display Reveals Allosteric Control of Ligand-Specific EGFR Activation. Acs Chemical Biology 2012, 7 (8), 1367-1376.

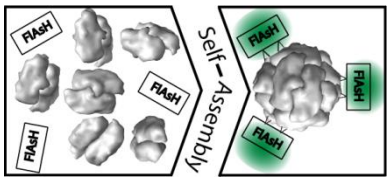
40. Enninga, J.; Mounier, J.; Sansonetti, P.; Van Nhieu, G. T., Secretion of type III effectors into host cells in real time. Nature Methods 2005, 2 (12), 959-965.

41. Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W., NIH Image to ImageJ: 25 years of image analysis. Nature Methods 2012, 9 (7), 671-675.

42. Kelly, S. M.; Jess, T. J.; Price, N. C., How to study proteins by circular dichroism. Biochimica Et Biophysica Acta-Proteins and Proteomics 2005, 1751 (2), 119-139.

43. Greenfield, N. J., Using circular dichroism spectra to estimate protein secondary structure. Nature Protocols 2006, 1 (6), 2876-2890.

44. Pace, C. N., Determination and analysis of urea and guanidine hydrochloride denaturation curves. Methods in enzymology 1986, 131, 266-80.


FlAsH
FlAsH
FlAsH
Self−Assembly
FlAsH
FlAsH
FlAsH

SUPPORTING INFORMATION

Detection of Specific Protein-Protein Interactions in Nanocages by Engineering Bipartite FlAsH Binding Sites

Thomas A. Cornell[†, Δ], Jing Fu[Δ, ‡], Stephanie H. Newland[#, ‡], and Brendan P. Orner[†, *].

[†] Department of Chemistry, King's College London, London, SE1 1DB, United Kingdom. [Δ] Division of Chemistry and Biological Chemistry, Nanyang Technological University, Singapore, 637371. [#] School of Chemistry, University of Southampton, Southampton, SO17 1BJ, United Kingdom.

‡These authors contributed equally.

Email - brendan.orner@kcl.ac.uk.

Table S1. Primers used for the construction of WTDps to be screened in lysates. Primers sourced from Supranom, Singapore.

Primer name	Primer sequence (5'-3')
Dps p22 F	GGA ACT CAT ATG AGC ACC GCA AAA CTG GTG
Dps p22 R	GGT ACT CTC GAT TTA TTC GAT GTT GCA TTC G

Table S2. Primers used for the construction of all designs to be used in lysates constructed via site directed mutagenesis. Primers sourced from Supranom, Singapore.

Primer name	Primer sequence (5'-3')
DpsCC F	GAA TGC AAC ATC GAA TGC TGC TGA TTC TCC TCA ACC ATG GCG ATA TC
DpsCC R	GAT ATC GCC ATG GTT GAG GAG AAT CAG CAG CAT TCG ATG TTG CAT TC
DpsGCC F	GAA TGC AAC ATC GAA GGA TGC TGC TGA TCC TCA ACC ATG GCG ATA TC
DpsGCC R	GAT ATC GCC ATG GTT GAG GAT CAG CAG CAT CCT TCG ATG TTG CAT TC
DpsGGCC	GAA TGC AAC ATC GAA GGA GGT TGC TGC TGA TCA ACC ATG GCG ATA TC
DpsGGCC	GAT ATC GCC ATG GTT GAT CAG CAG CAA CCT CCT TCG ATG TTG CAT TC
DpsGGGCC	GAA TGC AAC ATC GAA GGC GGT GGA TGC TGC TGA ACC ATG GCG ATA TC
DpsGGGCC R	GAT ATC GCC ATG GTT CAG CAG CAT CCA CCT CCT TCG ATG TTG CAT TC
DpsPAGCC F	GAA TGC AAC ATC GAA CCA GCG GGA TGC TGC TGA ACC ATG GCG ATA TC
DpsPAGCC R	GAT ATC GCC ATG GTT CAG CAA CAT CCC GCT GGT TCG ATG TTG CAT TC
DpsCCPGCC F	CAT CGA ACC AGC GGG ATG CTG CCC AGG TTG CTG CTA AAT GGC GAT ATC
DpsCCPGCC R	GAT ATC GCC ATT TAG CAG CAA CCT GGG CAG CAT CCC GCT GGT TCG ATG

Table S3. Primers used for the transfer of successful designs into pET-32b (for the production of purified proteins). Primers sourced from Integrated DNA technologies.

Primer name	Primer sequence (5'-3')
DpsGGGCC p32 R	GAG GAG AAG CCC GGT CAG CAG CAT CCA CCT CCT TC
DpsPAGCC p32 R	GAG GAG AAG CCC GGT CAG CAA CAT CCC GCT G
DpsCCPGCC p32 R	GAG GAG AAG CCC GGT TAG CAG CAA CCT GGG CAG
Dps F	GAC GAC GAC AAG ATG AGC ACC GCA AAA CTG GTG

Methods and materials S1 - Cloning of DpsCCPGCC, DpsCC, DpsGCC, DpsGGCC, DpsGGGCC, DpsPAGCC and Dps for Screening in Protein Lysates. The WT Dps gene was amplified with primers providing restriction sites. The PCR reaction (50 μl total volume) was a solution of reaction buffer (Fermentas, 5 μl of 10xPfu buffer), dNTP mix (Fermentas, 2 μl of a solution containing dATP, dTTP, dGTP, and dCTP 2 mM each), forward and reverse primers (400 ng each, Suprenom, Singapore), template (100 ng) and Pfu polymerase (Fermentas, 2 μl of 2.5 U/μl) which was subjected to an initial melting step at 95 ^{0}C for 30 s, followed by 30 cycles of amplification (95 ^{0}C for 30 s, 55 ^{0}C for 45 s and 72 ^{0}C for 1 min) followed by 72 ^{0}C for 7 min. The resulting PCR product was isolated by gel purification (Promega). The PCR product (300 ng) was digested (Tango buffer, Fermentas, 10 μl of 2x buffer), Nde1 (Fermentas, 1.5 μl of 20 U/μl), Xho1 (Fermentas, 1.5 μl of 20 U/μl), 50 μl total volume, 37 ^{0}C, 3 hours). The plasmid pET-22b (Novagen) was treated in the same way, followed by ligation of the insert with T4 ligase buffer (Fermentas, 2 μl, 10x buffer), digested insert (30 ng), digested plasmid (80 ng), T4 ligase (Fermentas, 0.2 μL of 20 U/μl), 20 μl total volume (37 ^{0}C, 2 hours). The constructs (2 μl) were transformed (Novablue, Novagen) and the resulting colonies were assessed by colony PCR, a solution of Gotaq buffer (Promega, 5 μl of 5x), dNTP mix (Promega, 2 μl of a solution containing dATP, dTTP, dGTP, and dCTP at 2 mM each), T7 promoter and terminator primers (150 ng each, Eurofins MWG), Gotaq DNA polymerase (Fermentas, 0.25 μl of 5 U/μl), colony suspension (10 μl of a 50 μl total colony suspension in deionised water) and deionised water (5 μl), was subjected to an initial melting step (95 ^{0}C for 10 min) followed by 30 cycles of amplification, ((95 ^{0}C for 30 s, 59 ^{0}C for 30 s and 72 ^{0}C for 30 s) followed by 72 ^{0}C for 7 min). Clones demonstrating a PCR product of the predicted size were subjected to miniprep (Invitrogen) and the resulting purified plasmids were sequenced (SI Figure S11 for sequencing results and SI Table S1 for primers).

Using the newly constructed **Dps** in pET-22b as a template, site directed mutagenesis was performed to create all other designs (Quikchange, Stratagene) by introducing mutations at the C-terminus. The reaction mixture containing Pfu buffer (Promega, 2 μl of 10x), dNTP mix (Promega, 2 μl of a solution containing dATP, dTTP, dGTP, and dCTP at 2 mM each), forward and reverse primers (125 ng each, Suprenom, Singapore), the dsDNA template (**Dps** pET-22b, 100 ng) and Pfu polymerase (Promega, 1 μl of 3 U/μl) in a total volume of 25 μl was subjected to melting (95 ^{0}C for 5 min), followed by 20 cycles of amplification (95 ^{0}C for 1 min, 55 ^{0}C for 1 min and 68 ^{0}C for 6 min) followed by 70 ^{0}C for 10 min. The reaction was then digested with Dpn1 (NEB buffer 4 (1 μl of 10x), amplification product (5 μl), deionized water (3.5 μl), Dpn1 (NEB, 0.7 μl of 10 U/μl), 37^{0}C, 2h). The resulting solution was transformed (XL-1 blue, Novagen) and the subsequent colonies subjected to miniprep (Promega, Wizard plus SV) and sequenced (SI Figure S11 for sequencing results and SI Table S2 for primers).

Methods and materials S2 - Cloning of all Designs for Screening as Purified Proteins. In order to express the proteins for isolation and purification, the genes were placed into a vector that provided affinity tags. The constructs were amplified from the mutated pET-22b plasmid with primers providing ligation independent cloning (LIC) sites. The PCR solution Pfu reaction buffer (Promega, 5 μl of 10x), dNTP mix (Fermentas, 2 μl of a solution containing dATP, dTTP, dGTP, and dCTP at 2 mM each), forward and reverse primers (400 ng each, Integrated DNA technologies), the template (100 ng) and Pfu polymerase (Promega, 2 μl of 2.5 U/μl) in 50 μl total volume was subjected to an initial melting step (95 ^{0}C for 30 s), followed by 30 cycles of amplification(95 ^{0}C for 30 s, 55 ^{0}C for 45 s and 72 ^{0}C for 1 min) followed by 72 ^{0}C for 7 min. The resulting PCR product was isolated by gel purification (Qiagen).The PCR product was then treated with T4 polymerase in order to create the required complementary overhangs (NEB buffer 2, 2 μl of 10x buffer), dATP (NEB, 2 μl of 25 mM), DTT (Sigma, 1 μl of 100 mM), BSA (NEB, 0.2 μl of 100x), PCR product (0.3 pmol) and T4 DNA polymerase (NEB, 0.6 μl of 10 U/μl) in a

total volume of 20 μl, 30 min, 22 ^{0}C). The vector, pET-32b (Novagen) was treated in the same way. The insert was annealed to open plasmid (1 μl of the T4 Pol product, 0.5 μl of plasmid solution (50 ng/μl) by incubating (room temperature, 30 min) followed by a second incubation (room temperature, 30 min) with an addition of EDTA (1 μl of 100 mM). The constructs (2.5 μl) were transformed (XL-1 Blue, Novagen) and the resulting colonies were assessed by colony PCR. A solution of Gotaq reaction buffer (5 μl of 5x), dNTP mix (Promega, 2 μl of a solution containing dATP, dTTP, dGTP, and dCTP at 2 mM each), T7 promoter and terminator primers (150 ng each, Eurofins MWG), Gotaq DNA polymerase (Promega, 0.25 μl of 5 U/μl), colony suspension (10 μl of a 50 μl total colony suspension in deionised water) and deionised water (5 μl), was subjected to an initial melting step (95 ^{0}C for 10 min) followed by amplification 30 cycles (95 ^{0}C for 30 s, 59 ^{0}C for 30 s and 72 ^{0}C for 30 s followed by 72 ^{0}C for 7 min). Clones demonstrating a PCR product of the predicted size were subjected to miniprep (Promega) and the resulting purified plasmid was sequenced (SI Figure S11 for sequence results and Table S3 for primers used).

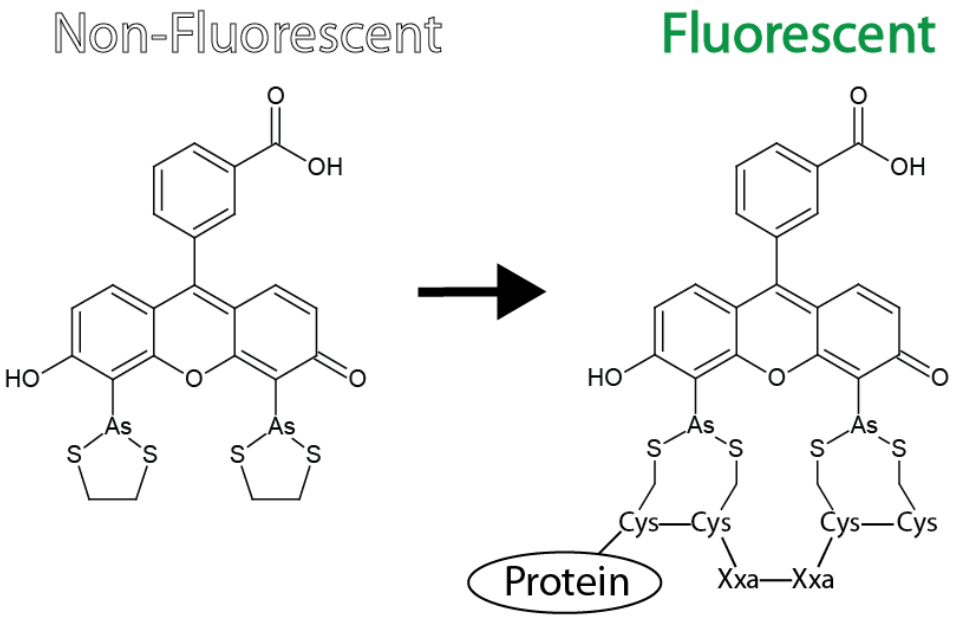


Figure S1 – FlAsH-EDT2 (left) quenched and non-fluorescent. Once EDT is exchanged with a tetracysteine tag, FlAsH becomes Fluorescent (right).

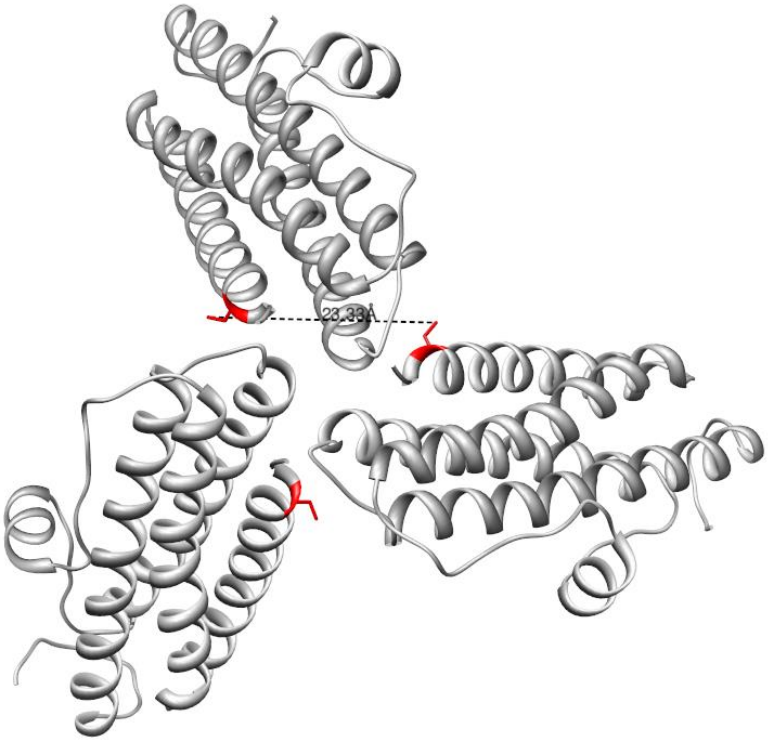


Figure S2 – Three fold axis of symmetry of Dps (PDB:1DPS) highlighting native cysteine residues in the structure (red). While there is only one per-structure and near the C-terminus, their distance apart is 23.33 Å (from sulfur to sulfur atom).

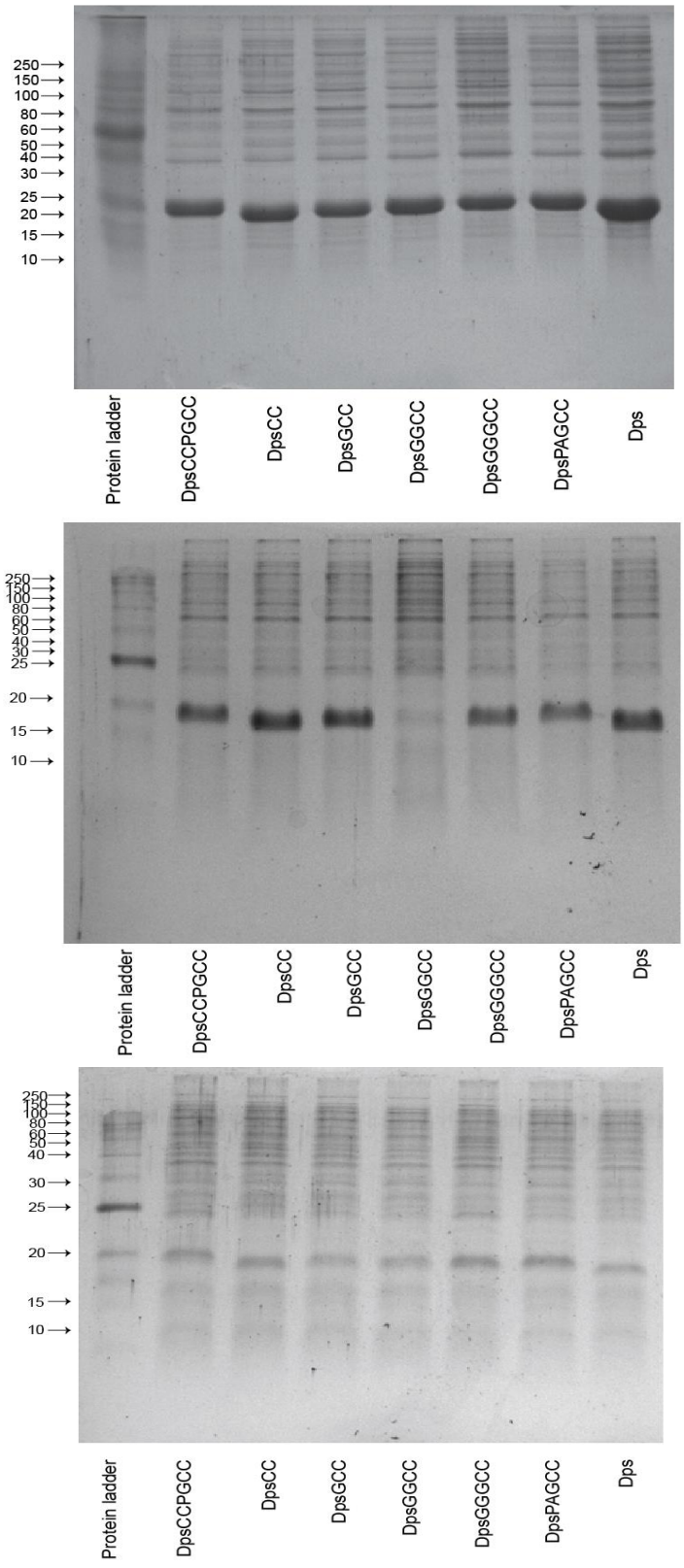


Figure S3 – SDS-PAGE (15 %) of lysate samples for each Dps design diluted to 1 mg/ml total protein concentration. Each design was expressed three times.

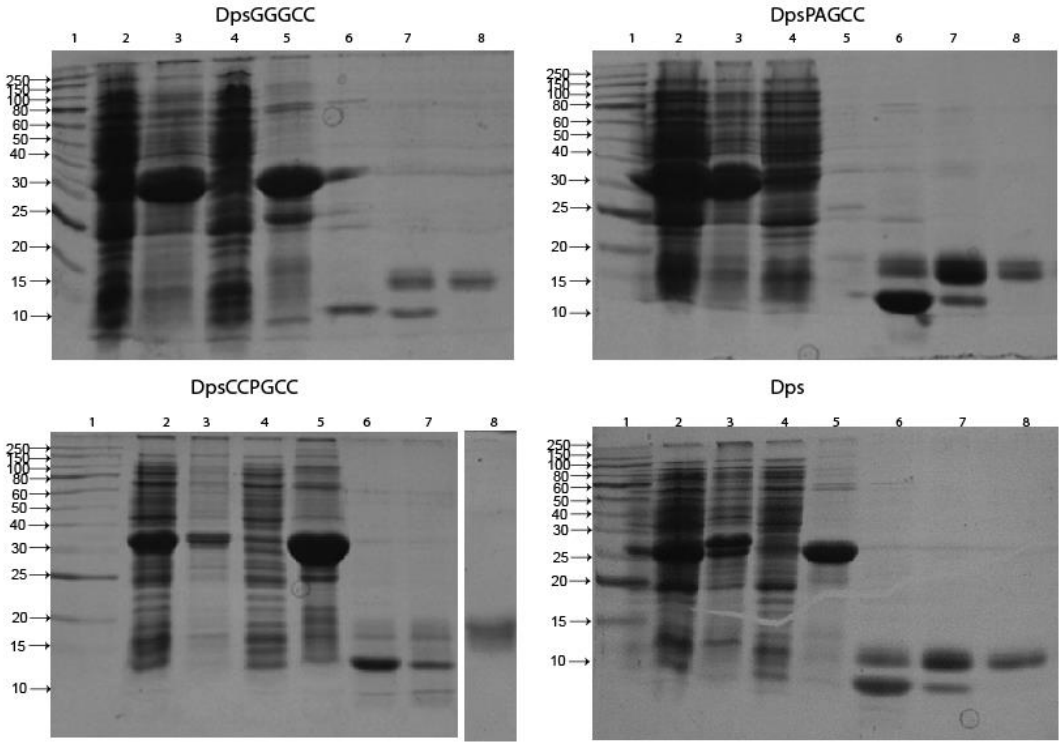


Figure S4 – SDS-PAGE gel (15 %) for each of the purified proteins. Lane 1: Protein standard ladder (NEB), Lane 2: Soluble fraction, lane 3: insoluble fraction, lane 4: wash from first histrap, lane 5: elution from first histrap, lane 6: elution from second histrap after EK digestion, lane 7: wash from second histrap after EK digestion, lane 8: SEC.

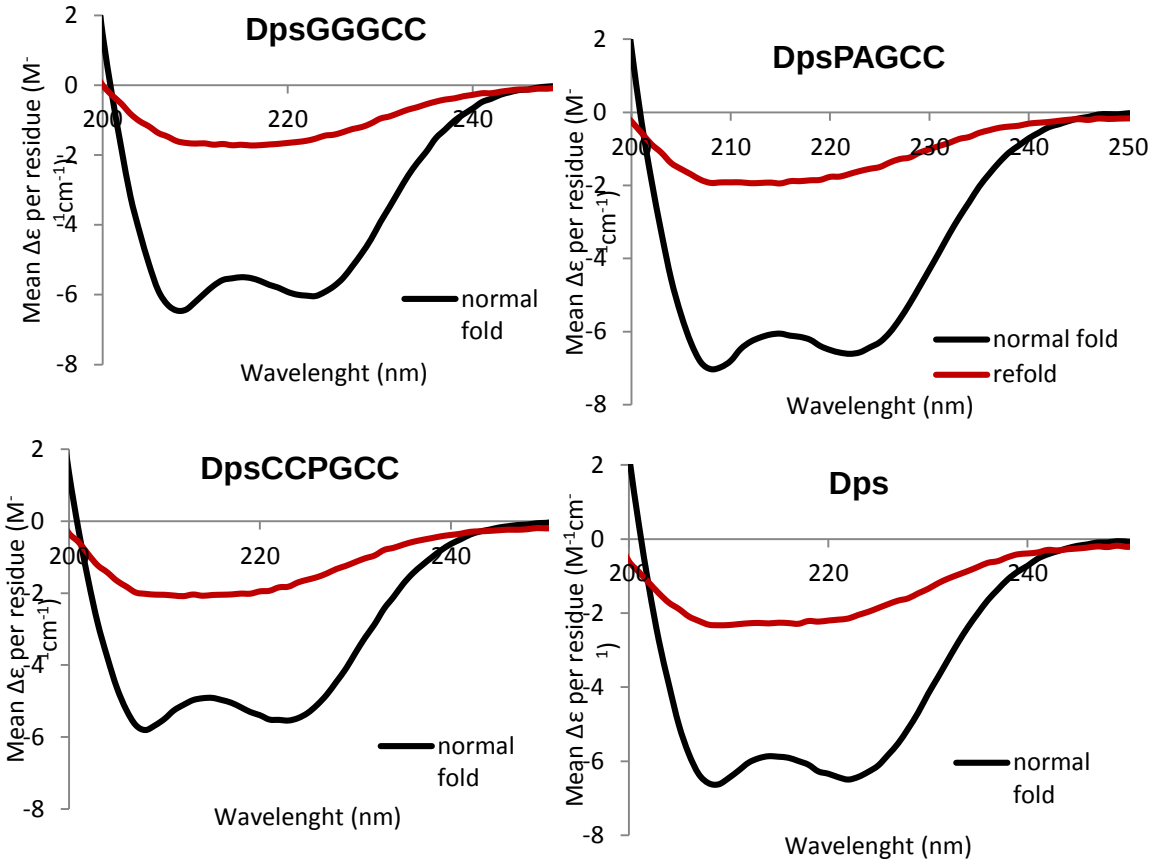


Figure S5 – Circular Dichroism spectra demonstrating the effect of heating on the fold of each protein. Black: spectrum before denaturation. Red: spectrum after slow cooling post-denaturation.

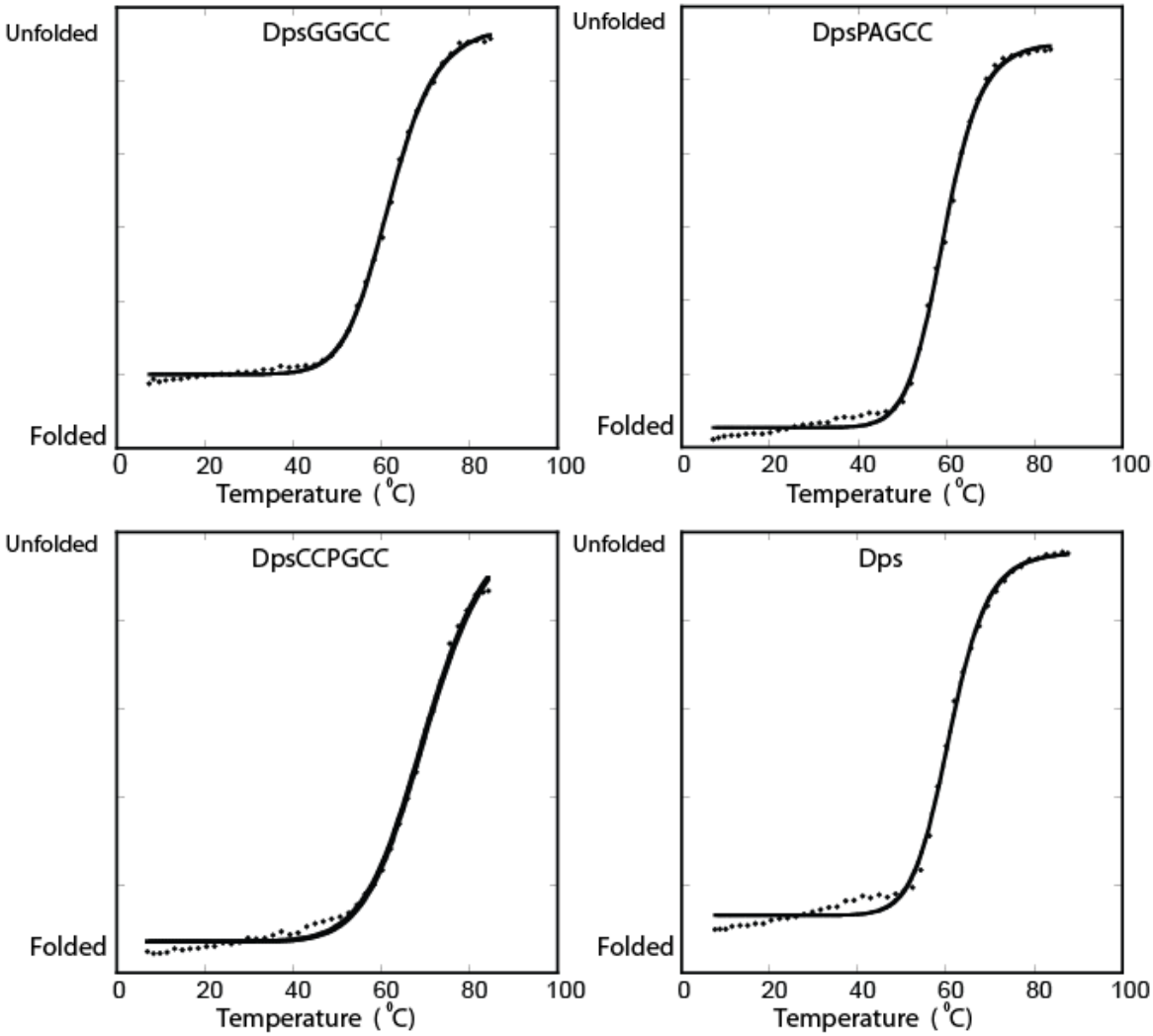


Figure S6 – Thermal melting plots of Dps designs and controls monitored at 222 nm in CD. Dotted lines indicate fit to a two state unfolding (m1+(m2-m1)/(1+(m0/m3)^m4)[equation.1, 2, 3]

Table S4 – Melting temperature of all repeats from thermal denaturation. S.D from three replicates.

Design	Tm average (°C)
DpsCCPGCC	70.35 ± 1.42
DpsGGGCC	61.87 ± 0.13
DpsPAGCC	58.84 ± 0.40
Dps	61.66 ± 1.49

Table S5 – Electrospray mass spectrometry from purified protein samples.

Protein	Measured mass (Da)	Expected mass (Da)	Difference (Da)
DpsCCPGCC	19273.0	19278.0	5.0
DpsGGGCC	19100.3	19088.8	-11.5
DpsPAGCC	19154.0	19142.8	-11.2
Dps	18710.2	18711.3	1.1

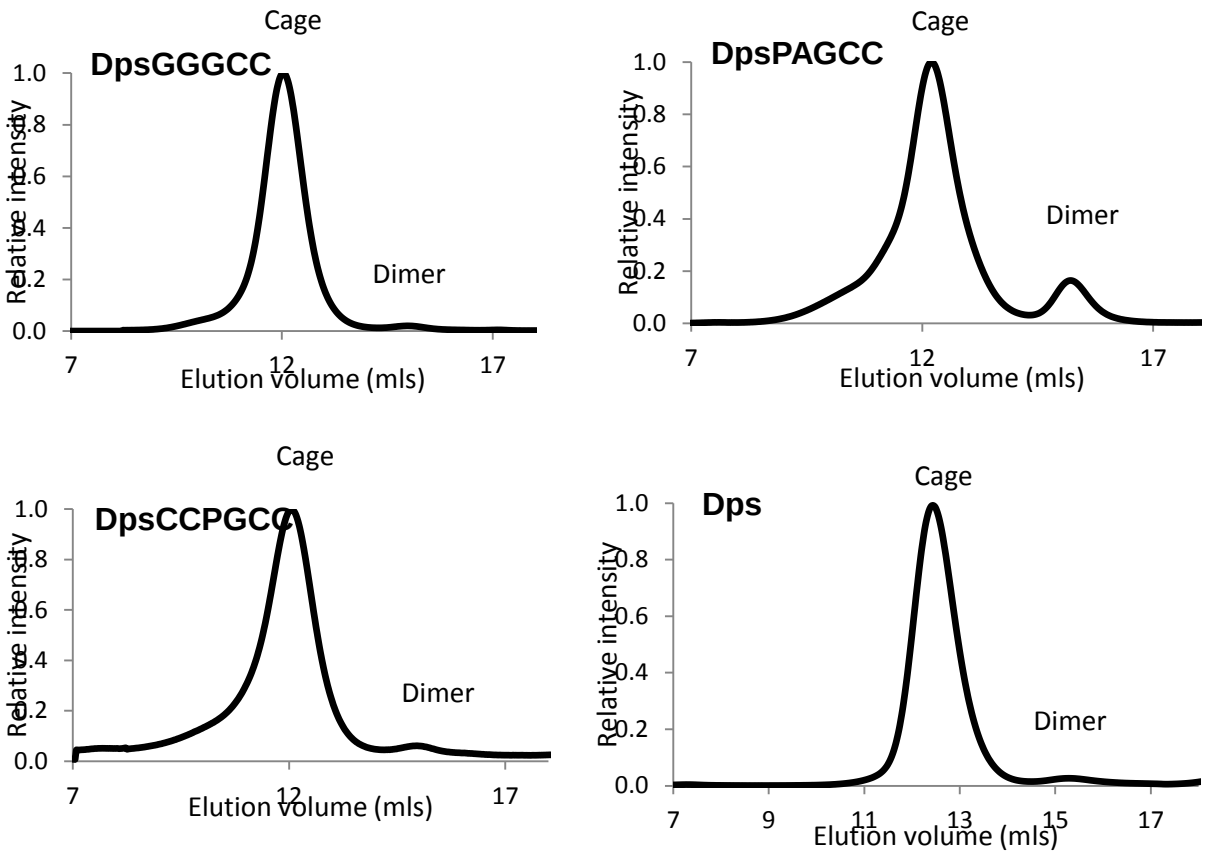
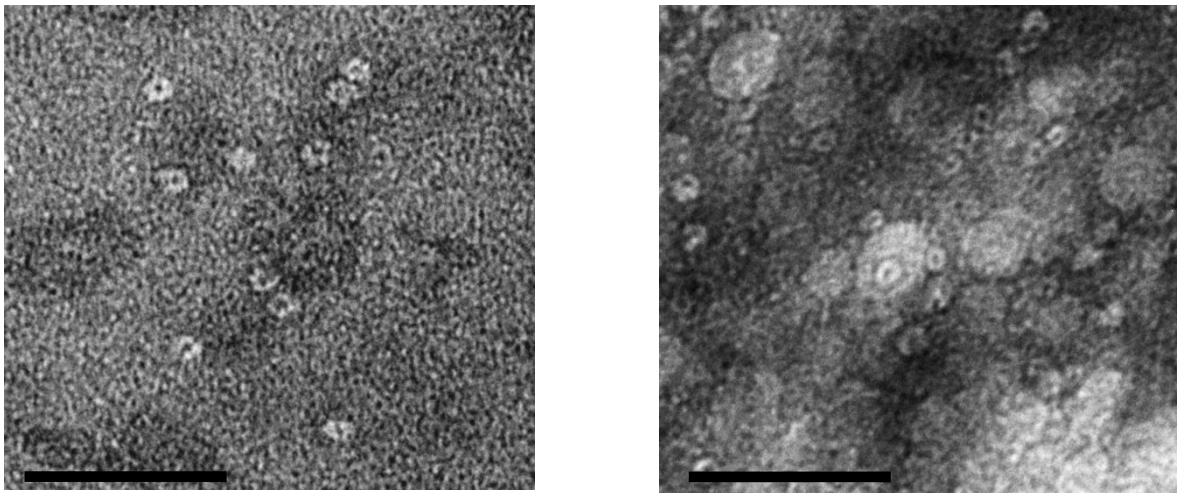
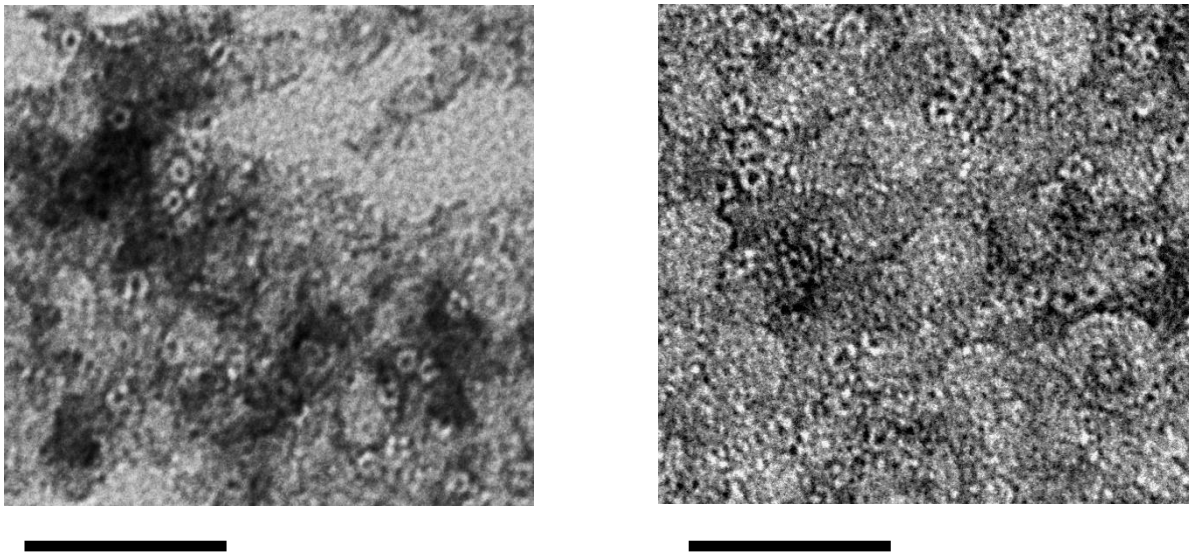


Figure S7. Analytical size exclusion chromatography data for all purified Dps designs and controls. Each was repeated three times using a 0.5 ml injection of 1 mg/ml protein in a GE Superdex 200pg column, with the averages shown above.

DpsGGGCC



DpsPAGCC



DpsCCPGCC

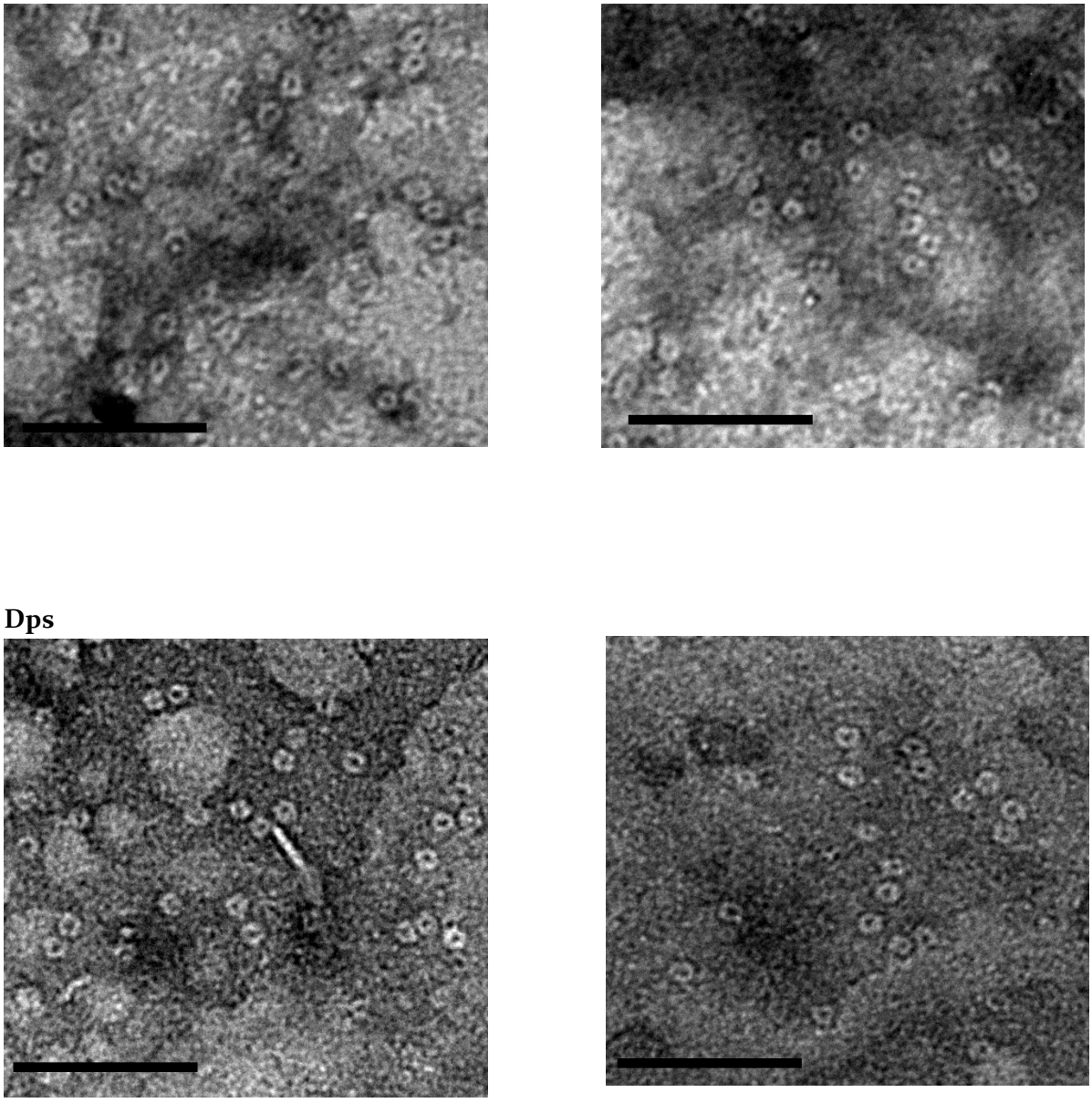


Figure S8 – TEM micrographs for characterization of DpsGGGCC, DpsPAGCC, DpsCCPGCC and Dps. Scale bar is equal to 50 nm.

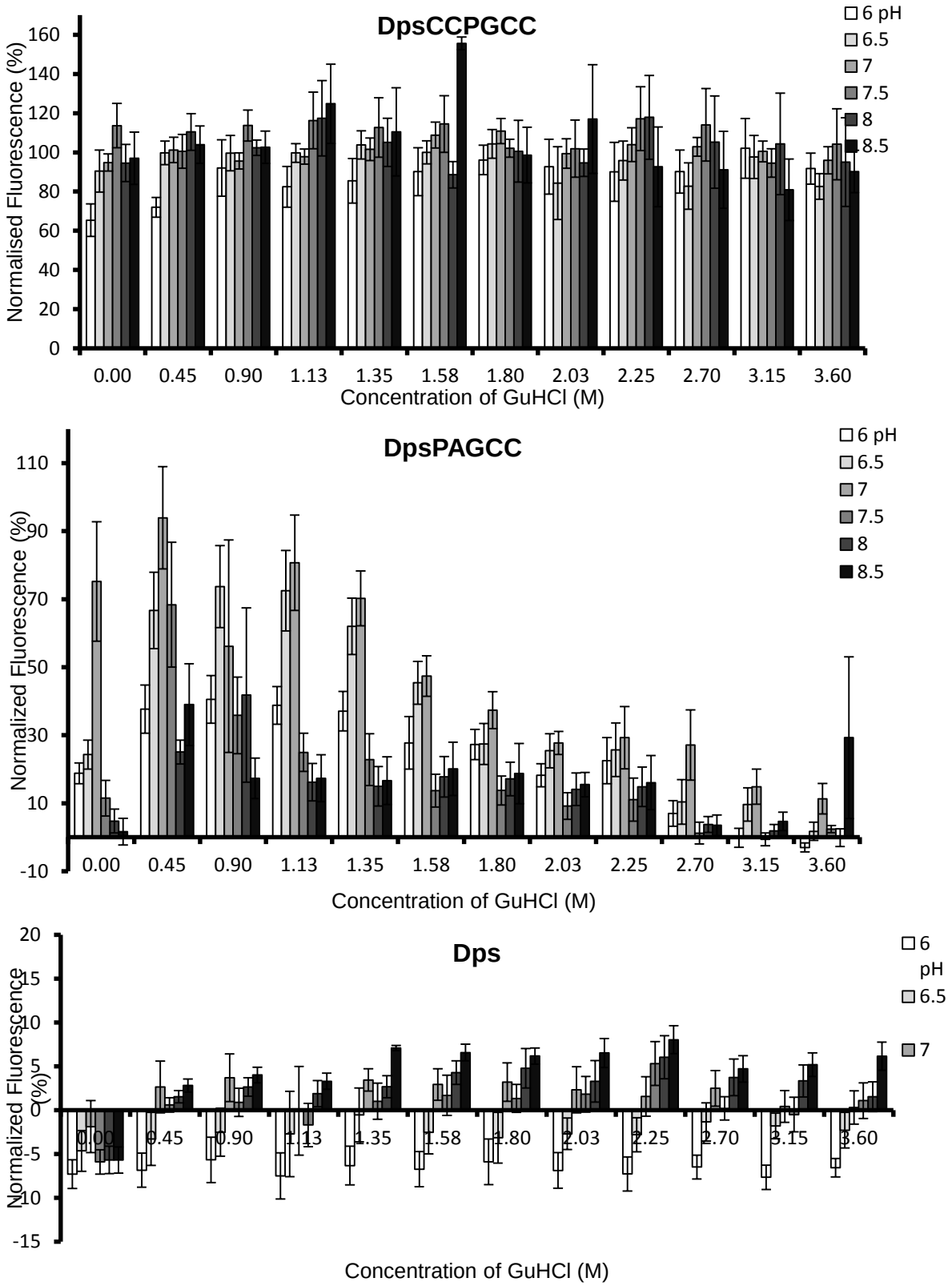


Figure S9 –Raw data for pH vs GuHCl (See Figure 4b). Showing the effect on the stability of DpsPAGCC at different pH's by monitoring the fluorescence. Data normalized using the controls DpsCCPGCC and Dps by taking the average reading as 100 and zero percent respectively.

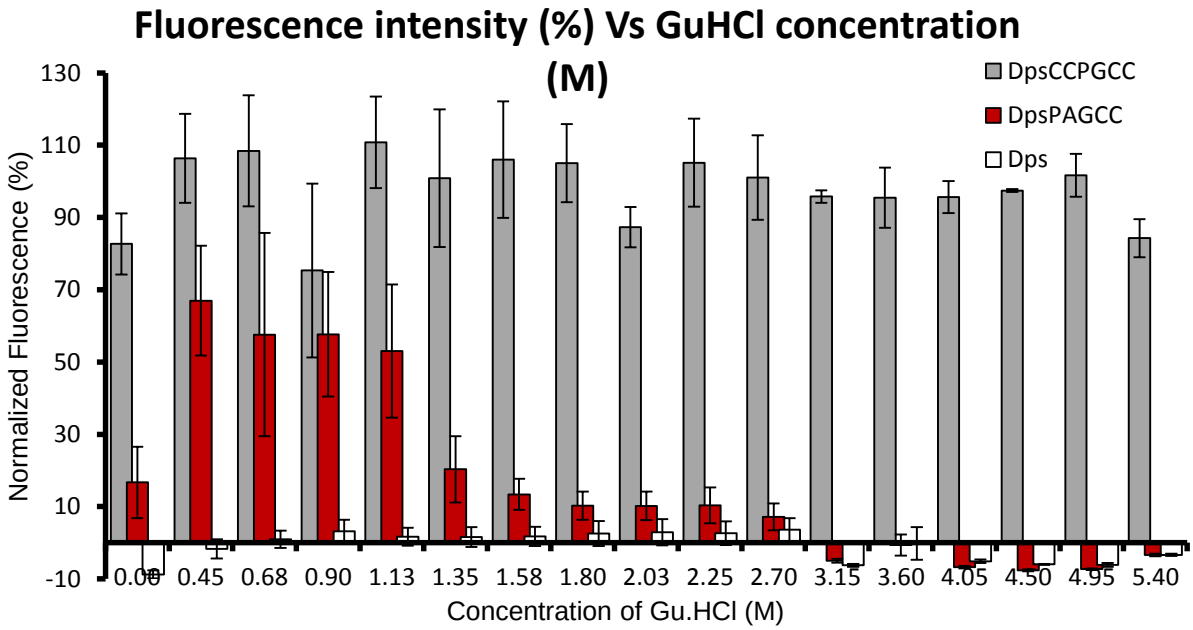


Figure S10 – Monitoring fluorescence loss during protein unfolding for DpsPAGCC at pH 7.8. (See Figure 4a) Data normalized using the controls DpsCCPGCC and Dps by taking the average reading as 100 and zero percent respectively.

Dps in pET-22b

```
Query  50   ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  109
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  60

Query  110  GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  169
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  170  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  229
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  230  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  289
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  290  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  349
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  350  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  409
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  410  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  469
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  420

Query  470  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  529
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  530  TTCATCGAATGCAACATCGAATAA  553
            ||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAATAA  504
```

DpsCCPGCC in pET-22b

```
Query  41   ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  60

Query  101  GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  161  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  221  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  281  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  341  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  400
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  401  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  460
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  420

Query  461  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  520
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  521  TTCATCGAATGCAACATCGAATGCTGCCCAGGTTGCTGCTAA  562
            ||||||||||||||||||||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAATGCTGCCCAGGTTGCTGCTAA  522
```

DpsCC in pET-22b

```
Query  30   ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  89
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  60

Query  90   GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  149
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  150  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  209
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  210  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  269
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  270  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  329
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  330  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  389
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  390  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  449
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  420

Query  450  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  509
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  510  TTCATCGAATGCAACATCGAATGCTGCTGA  539
            ||||||||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAATGCTGCTGA  510
```

DpsGCC in pET-22b

```
Query  14   ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  73
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  60

Query  74   GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  133
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  134  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  193
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  194  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  253
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  254  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  313
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  314  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  373
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  374  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  433
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  420

Query  434  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  493
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  494  TTCATCGAATGCAACATCGAAGGATGCTGCTGA  526
            ||||||||||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAAGGATGCTGCTGA  513
```

DpsGGCC in pET-22b

```
Query  40   ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  99
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  60

Query  100  GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  159
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  160  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  219
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  220  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  279
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  280  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  339
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  340  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  399
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  400  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  459
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  420

Query  460  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  519
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  520  TTCATCGAATGCAACATCGAAGGAGGTTGCTGCTGA  555
            ||||||||||||||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAAGGAGGTTGCTGCTGA  516
```

DpsGGGCC in pET-22b

```
Query  37   ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  96
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  60
```

```
Query  97   GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  156
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  157  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  216
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  217  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  276
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  277  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  336
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  337  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  396
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  397  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  456
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  420

Query  457  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  516
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  517  TTCATCGAATGCAACATCGAAGGAGGTGGATGCTGCTGA  555
            ||||||||||||||||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAAGGAGGTGGATGCTGCTGA  519
```

DpsPAGCC in pET-22b

```
Query  39   ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  98
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  60

Query  99   GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  158
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  159  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  218
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  219  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  278
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  279  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  338
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  339  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  398
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  399  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  458
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  420

Query  459  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  518
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  519  TTCATCGAATGCAACATCGAACCAGCGGGATGTTGCTGA  557
```

```
            ||||||||||||||||||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAACCAGCGGGATGTTGCTGA  519
```

Dps in pET-32b

```
Query  9    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  68
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  60

Query  69   GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  128
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  129  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  188
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  189  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  248
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  249  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  308
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  309  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  368
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  369  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  428
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  420

Query  429  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  488
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  489  TTCATCGAATGCAACATCGAATAA  512
            ||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAATAA  504
```

DpsCCPGCC in pET-32b

```
Query  9    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  68
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  60

Query  69   GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  128
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  129  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  188
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  189  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  248
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  249  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  308
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  309  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  368
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360
```

```
Query  369  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  428
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  420

Query  429  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  488
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  489  TTCATCGAATGCAACATCGAATGCTGCCCAGGTTGCTGCTAA  530
            ||||||||||||||||||||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAATGCTGCCCAGGTTGCTGCTAA  522
```

DpsGGGCC in pET-32b

```
Query  9    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  68
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  60

Query  69   GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  128
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  129  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  188
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  189  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  248
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240

Query  249  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  308
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241  GCAGAACGTGCAGTGCAGCTGGGTGGTGTGGCACTGGGTACCACCCAGGTGATCAACAGC  300

Query  309  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  368
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301  AAAACCCCGCTGAAAAGCTACCCGCTGGATATCCATAACGTGCAGGATCATCTGAAAGAA  360

Query  369  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  428
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  CTGGCAGATCGTTACGCAATCGTGGCAAACGATGTGCGTAAAGCAATCGGTGAAGCAAAA  420

Query  429  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  488
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421  GATGATGATACCGCAGATATCCTGACCGCAGCAAGCCGTGATCTGGATAAATTCCTGTGG  480

Query  489  TTCATCGAATGCAACATCGAAGGAGGTGGATGCTGCTGA  527
            ||||||||||||||||||||||||||||||||||||||||
Sbjct  481  TTCATCGAATGCAACATCGAAGGAGGTGGATGCTGCTGA  519
```

DpsPAGCC in pET-32b

```
Query  11   ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  70
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1    ATGAGCACCGCAAAACTGGTGAAAAGCAAAGCAACCAACCTGCTGTACACCCGTAACGAT  60

Query  71   GTGAGCGATAGCGaaaaaaaaGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  130
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  61   GTGAGCGATAGCGAAAAAAAAAGCAACCGTGGAACTGCTGAACCGTCAGGTGATCCAGTTC  120

Query  131  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  190
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  ATCGATCTGAGCCTGATCACCAAACAGGCACATTGGAACATGCGTGGTGCAAACTTCATC  180

Query  191  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  250
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181  GCAGTGCATGAAATGCTGGATGGTTTCCGTACCGCACTGATCGATCATCTGGATACCATG  240
```

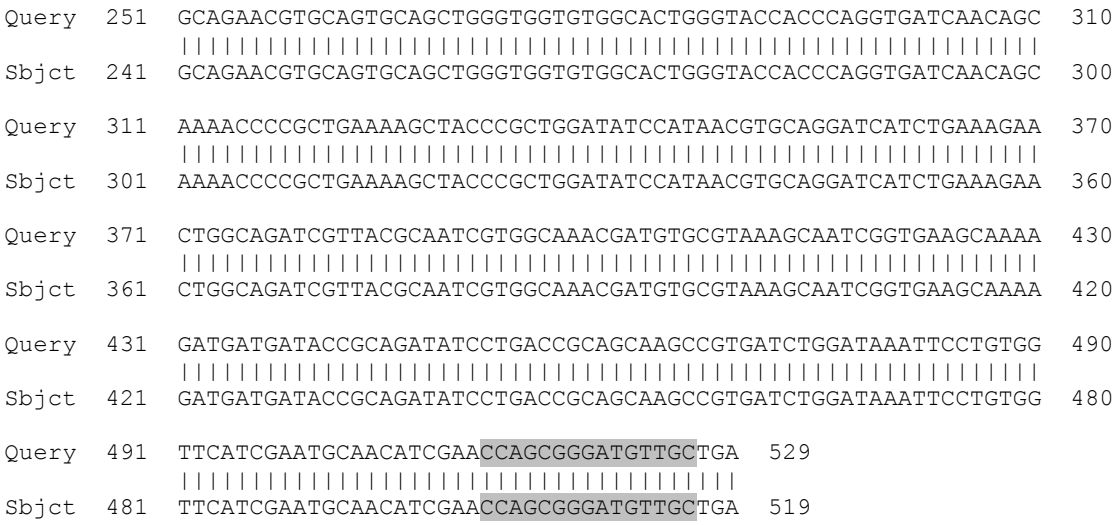


Figure S11. Comparison of the sequencing results with expected sequence. Sequencing result shown on the top line, aligned with the expected sequence on the bottom. Yellow indicates extension or codon mutation. Protein design and vector used is indicated in each sequence alignment title.

REFERENCES

1. Fan, R.; Boyle, A. L.; Cheong, V. V.; Ng, S. L.; Orner, B. P., A Helix Swapping Study of Two Protein Cages. *Biochemistry* **2009,** *48* (24), 5623-5630.
2. Zhang, Y.; Fu, J.; Chee, S. Y.; Ang, E. X. W.; Orner, B. P., Rational disruption of the oligomerization of the mini-ferritin E. coli DPS through protein-protein interface mutation. *Protein Science* **2011,** *20* (11), 1907-1917.
3. Zhang, Y.; Orner, B. P., Self-Assembly in the Ferritin Nano-Cage Protein Superfamily. *International Journal of Molecular Sciences* **2011,** *12* (8), 5406-5421.

Total list of references

1. Whitesides, G. M.; Grzybowski, B., Self-assembly at all scales. *Science* **2002,** *295* (5564), 2418-2421.
2. Costerton, J. W.; Cheng, K. J.; Geesey, G. G.; Ladd, T. I.; Nickel, J. C.; Dasgupta, M.; Marrie, T. J., Bacterial biofilms in nature and disease. *Annual Review of Microbiology* **1987,** *41*, 435-464.
3. Zhang, S. G., Fabrication of novel biomaterials through molecular self-assembly. *Nature Biotechnology* **2003,** *21* (10), 1171-1178.
4. Janin, J.; Bahadur, R. P.; Chakrabarti, P., Protein-protein interaction and quaternary structure. *Quarterly Reviews of Biophysics* **2008,** *41* (2), 133-180.
5. Dill, K. A.; Ozkan, S. B.; Shell, M. S.; Weikl, T. R., The protein folding problem. In *Annual Review of Biophysics*, 2008; Vol. 37, pp 289-316.
6. Walter, S.; Buchner, J., Molecular chaperones - Cellular machines for protein folding. *Angewandte Chemie-International Edition* **2002,** *41* (7), 1098-1113.
7. Bromley, E. H. C.; Channon, K.; Moutevelis, E.; Woolfson, D. N., Peptide and protein building blocks for synthetic biology: From programming biomolecules to self-organized biomolecular systems. *Acs Chemical Biology* **2008,** *3* (1), 38-50.
8. Pawson, T.; Nash, P., Protein-protein interactions define specificity in signal transduction. *Genes & Development* **2000,** *14* (9), 1027-1047.
9. Prives, C.; Hall, P. A., The P53 pathway. *Journal of Pathology* **1999,** *187* (1), 112-126.
10. Dalrymple, B. P.; Kongsuwan, K.; Wijffels, G.; Dixon, N. E.; Jennings, P. A., A universal protein-protein interaction motif in the eubacterial DNA replication and repair systems. *Proceedings of the National Academy of Sciences of the United States of America* **2001,** *98* (20), 11627-11632.
11. Halliburton, W. D., On Muscle-Plasma. *The Journal of physiology* **1887,** *8* (3-4), 133-202.
12. Pollard, T. D.; Borisy, G. G., Cellular motility driven by assembly and disassembly of actin filaments. *Cell* **2003,** *112* (4), 453-465.
13. Pellegrini, F.; Budman, D. R., Review: Tubulin function, action of antitubulin drugs, and new drug development. *Cancer Investigation* **2005,** *23* (3), 264-273.
14. Selkoe, D. J., Alzheimer's disease: Genes, proteins, and therapy. *Physiological Reviews* **2001,** *81* (2), 741-766.
15. Vivekanandan, S.; Brender, J. R.; Lee, S. Y.; Ramamoorthy, A., A partially folded structure of amyloid-beta(1-40) in an aqueous environment. *Biochemical and Biophysical Research Communications* **2011,** *411* (2), 312-316.
16. Prusiner, S. B., Prions. *Proceedings of the National Academy of Sciences of the United States of America* **1998,** *95* (23), 13363-13383.
17. Bogan, A. A.; Thorn, K. S., Anatomy of hot spots in protein interfaces. *Journal of Molecular Biology* **1998,** *280* (1), 1-9.
18. Ellis, J., Porcine Circovirus A Historical Perspective. *Veterinary Pathology* **2014,** *51* (2), 315-327.
19. Legendre, M.; Bartoli, J.; Shmakova, L.; Jeudy, S.; Labadie, K.; Adrait, A.; Lescot, M.; Poirot, O.; Bertaux, L.; Bruley, C.; Coute, Y.; Rivkina, E.; Abergel, C.; Claverie, J.-M., Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. *Proceedings of the National Academy of Sciences of the United States of America* **2014,** *111* (11), 4274-4279.
20. Zandi, R.; Reguera, D.; Bruinsma, R. F.; Gelbart, W. M.; Rudnick, J., Origin of icosahedral symmetry in viruses. *Proceedings of the National Academy of Sciences of the United States of America* **2004,** *101* (44), 15556-15560.

21. Burnett, R. M., The structure of the adenovirus capsid .2. the packing symmetry of hexon and its implications for viral architecture. *Journal of Molecular Biology* **1985,** *185* (1), 125-143.
22. Yamada, S.; Matsuzawa, T.; Yamada, K.; Yoshioka, S.; Ono, S.; Hishinuma, T., Modified inversion recovery method for NMR imaging *Science Reports of the Research Institutes Tohoku University Series C Medicine* **1986,** *33* (1-4), 9-15.
23. Ohtsuka, K.; Hata, M., Molecular chaperone function of mammalian Hsp70 and Hsp40 - a review. *International Journal of Hyperthermia* **2000,** *16* (3), 231-245.
24. Zhang, X.; Konarev, P. V.; Petoukhov, M. V.; Svergun, D. I.; Xing, L.; Cheng, R. H.; Haase, I.; Fischer, M.; Bacher, A.; Ladenstein, R.; Meining, W., Multiple assembly states of lumazine synthase: A model relating catalytic function and molecular assembly. *Journal of Molecular Biology* **2006,** *362* (4), 753-770.
25. Kedersha, N. L.; Heuser, J. E.; Chugani, D. C.; Rome, L. H., Vaults .3. Vault ribonucleoprotein-particles open into flower-like structures with octagonal symmetry. *Journal of Cell Biology* **1991,** *112* (2), 225-235.
26. Tanaka, H.; Kato, K.; Yamashita, E.; Sumizawa, T.; Zhou, Y.; Yao, M.; Iwasaki, K.; Yoshimura, M.; Tsukihara, T., The Structure of Rat Liver Vault at 3.5 Angstrom Resolution. *Science* **2009,** *323* (5912), 384-388.
27. Dautant, A.; Meyer, J. B.; Yariv, J.; Precigoux, G.; Sweet, R. M.; Kalb, A. J.; Frolow, F., Structure of a monoclinic crystal form of cytochrome b1 (bacterioferritin) from E-coli. *Acta Crystallographica Section D-Biological Crystallography* **1998,** *54*, 16-24.
28. Zhang, X. F.; Meining, W.; Fischer, M.; Bacher, A.; Ladenstein, R., X-ray structure analysis and crystallographic refinement of lumazine synthase from the hyperthermophile Aquifex aeolicus at 1.6 angstrom resolution: Determinants of thermostability revealed from structural comparisons. *Journal of Molecular Biology* **2001,** *306* (5), 1099-1114.
29. Kaufmann, B.; Simpson, A. A.; Rossmann, M. G., The structure of human parvovirus B19. *Proceedings of the National Academy of Sciences of the United States of America* **2004,** *101* (32), 11628-11633.
30. Vu, N. T.; Moriwaki, Y.; Caaveiro, J. M. M.; Terada, T.; Tsutsumi, H.; Hamachi, I.; Shimizu, K.; Tsumoto, K., Selective binding of antimicrobial porphyrins to the heme-receptor IsdH-NEAT3 of Staphylococcus aureus. *Protein Science* **2013,** *22* (7), 942-953.
31. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF chimera - A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **2004,** *25* (13), 1605-1612.
32. Sapsford, K. E.; Algar, W. R.; Berti, L.; Gemmill, K. B.; Casey, B. J.; Oh, E.; Stewart, M. H.; Medintz, I. L., Functionalizing Nanoparticles with Biological Molecules: Developing Chemistries that Facilitate Nanotechnology. *Chemical Reviews* **2013,** *113* (3), 1904-2074.
33. Kitagawa, T.; Kosuge, H.; Uchida, M.; Dua, M. M.; Iida, Y.; Dalman, R. L.; Douglas, T.; McConnell, M. V., RGD-Conjugated Human Ferritin Nanoparticles for Imaging Vascular Inflammation and Angiogenesis in Experimental Carotid and Aortic Disease. *Molecular Imaging and Biology* **2012,** *14* (3), 315-324.
34. Li, K.; Zhang, Z.-P.; Luo, M.; Yu, X.; Han, Y.; Wei, H.-P.; Cui, Z.-Q.; Zhang, X.-E., Multifunctional ferritin cage nanostructures for fluorescence and MR imaging of tumor cells. *Nanoscale* **2012,** *4* (1), 188-193.
35. Douglas, T.; Young, M., Viruses: Making friends with old foes. *Science* **2006,** *312* (5775), 873-875.
36. Theil, E. C., Ferritin - structure, gene-regulation, and cellular function in animals, plants, and microorganisms. *Annual Review of Biochemistry* **1987,** *56*, 289-315.

37. Hoare, R. J.; Harrison, P. M.; Hoy, T. G., Structure of horse-spleen apoferritin at 6-A resolution. *Nature* **1975,** *255* (5510), 653-654.
38. Crichton, R. R.; Declercq, J.-P., X-ray structures of ferritins and related proteins. *Biochimica Et Biophysica Acta-General Subjects* **2010,** *1800* (8), 706-718.
39. Andrews, S. C.; Arosio, P.; Bottke, W.; Briat, J. F.; Vondarl, M.; Harrison, P. M.; Laulhere, J. P.; Levi, S.; Lobreaux, S.; Yewdall, S. J., Structure, function, and evolution of ferritins. *Journal of Inorganic Biochemistry* **1992,** *47* (3-4), 161-174.
40. Fan, R.; Boyle, A. L.; Cheong, V. V.; Ng, S. L.; Orner, B. P., A Helix Swapping Study of Two Protein Cages. *Biochemistry* **2009,** *48* (24), 5623-5630.
41. Ardejani, M. S.; Chok, X. L.; Foo, C. J.; Orner, B. P., Complete shift of ferritin oligomerization toward nanocage assembly via engineered protein-protein interactions. *Chemical Communications* **2013,** *49* (34), 3528-3530.
42. Johnson, E.; Cascio, D.; Sawaya, M. R.; Gingery, M.; Schroder, I., Crystal structures of a tetrahedral open pore ferritin from the hyperthermophilic Archaeon Archaeoglobus fulgidus. *Structure* **2005,** *13* (4), 637-648.
43. Aisen, P.; Listowsky, I., Iron transport and storage proteins. *Annual Review of Biochemistry* **1980,** *49*, 357-393.
44. Yariv, J.; Kalb, A. J.; Sperling, R.; Bauminger, E. R.; Cohen, S. G.; Ofer, S., The composition and the structure of bacterioferritin of Escherichia-coli. *Biochemical Journal* **1981,** *197* (1), 171-175.
45. Hintze, K. J.; Theil, E. C., Cellular regulation and molecular interactions of the ferritins. *Cellular and Molecular Life Sciences* **2006,** *63* (5), 591-600.
46. Zhang, Y.; Raudah, S.; Teo, H.; Teo, G. W. S.; Fan, R.; Sun, X.; Orner, B. P., Alanine-shaving Mutagenesis to Determine Key Interfacial Residues Governing the Assembly of a Nano-cage Maxi-ferritin. *Journal of Biological Chemistry* **2010,** *285* (16), 12078-12086.
47. Ardejani, M. S.; Li, N. X.; Orner, B. P., Stabilization of a Protein Nanocage through the Plugging of a Protein-Protein Interfacial Water Pocket. *Biochemistry* **2011,** *50* (19), 4029-4037.
48. Ilari, A.; Stefanini, S.; Chiancone, E.; Tsernoglou, D., The dodecameric ferritin from Listeria innocua contains a novel intersubunit iron-binding site. *Nature Structural Biology* **2000,** *7* (1), 38-43.
49. Chiancone, E.; Ceci, P., The multifaceted capacity of Dps proteins to combat bacterial stress conditions: Detoxification of iron and hydrogen peroxide and DNA binding. *Biochimica Et Biophysica Acta-General Subjects* **2010,** *1800* (8), 798-805.
50. Zhang, Y.; Fu, J.; Chee, S. Y.; Ang, E. X. W.; Orner, B. P., Rational disruption of the oligomerization of the mini-ferritin E. coli DPS through protein-protein interface mutation. *Protein Science* **2011,** *20* (11), 1907-1917.
51. Grant, R. A.; Filman, D. J.; Finkel, S. E.; Kolter, R.; Hogle, J. M., The crystal structure of Dps, a ferritin homolog that binds and protects DNA. *Nature Structural Biology* **1998,** *5* (4), 294-303.
52. Theil, E. C., Coordinating responses to iron and oxygen stress with DNA and mRNA promoters: The ferritin story. *Biometals* **2007,** *20* (3-4), 513-521.
53. Chiancone, E.; Ceci, P.; Ilari, A.; Ribacchi, F.; Stefanini, S., Iron and proteins for iron storage and detoxification. *Biometals* **2004,** *17* (3), 197-202.
54. Bevers, L. E.; Theil, E. C., Maxi- and mini-ferritins: minerals and protein nanocages. *Progress in molecular and subcellular biology* **2011,** *52*, 29-47.
55. Li, L.; Fang, C. J.; Ryan, J. C.; Niemi, E. C.; Lebron, J. A.; Bjorkman, P. J.; Arase, H.; Torti, F. M.; Torti, S. V.; Nakamura, M. C.; Seaman, W. E., Binding and uptake of H-ferritin are mediated by human transferrin receptor-1. *Proceedings of the National Academy of Sciences of the United States of America* **2010,** *107* (8), 3505-3510.

56. Lim, S.; Ieee, Protein Cages as Theranostic Agent Carriers. *6th Ieee International Conference on Nano/Molecular Medicine and Engineering (Ieee-Nanomed 2012)* **2012**.
57. Harrison, P. M.; Arosio, P., Ferritins: Molecular properties, iron storage function and cellular regulation. *Biochimica Et Biophysica Acta-Bioenergetics* **1996,** *1275* (3), 161-203.
58. Baraibar, M. A.; Barbeito, A. G.; Muhoberac, B. B.; Vidal, R., Iron-mediated Aggregation and a Localized Structural Change Characterize Ferritin from a Mutant Light Chain Polypeptide That Causes Neurodegeneration. *Journal of Biological Chemistry* **2008,** *283* (46), 31679-31689.
59. Curtis, A. R. J.; Fey, C.; Morris, C. M.; Bindoff, L. A.; Ince, P. G.; Chinnery, P. F.; Coulthard, A.; Jackson, M. J.; Jackson, A. P.; McHale, D. P.; Hay, D.; Barker, W. A.; Markham, A. F.; Bates, D.; Curtis, A.; Burn, J., Mutation in the gene encoding ferritin light polypeptide causes dominant adult-onset basal ganglia disease. *Nature Genetics* **2001,** *28* (4), 350-354.
60. Butts, C. A.; Swift, J.; Kang, S. G.; Di Costanzo, L.; Christiansen, D. W.; Saven, J. G.; Dmochowski, I. J., Directing noble metal ion chemistry within a designed ferritin protein. *Biochemistry* **2008,** *47* (48), 12729-12739.
61. Zhang, L.; Swift, J.; Butts, C. A.; Yerubandi, V.; Dmochowski, I. J., Structure and activity of apoferritin-stabilized gold nanoparticles. *Journal of Inorganic Biochemistry* **2007,** *101* (11-12), 1719-1729.
62. Fan, R.; Chew, S. W.; Cheong, V. V.; Orner, B. P., Fabrication of Gold Nanoparticles Inside Unmodified Horse Spleen Apoferritin. *Small* **2010,** *6* (14), 1483-1487.
63. Keyes, J. D.; Hilton, R. J.; Farrer, J.; Watt, R. K., Ferritin as a photocatalyst and scaffold for gold nanoparticle synthesis. *Journal of Nanoparticle Research* **2011,** *13* (6), 2563-2575.
64. Butts, C. A.; Swift, J.; Kang, S.-g.; Di Costanzo, L.; Christianson, D. W.; Saven, J. G.; Dmochowski, I. J., Directing Noble Metal Ion Chemistry within a Designed Ferritin Protein. *Biochemistry* **2008,** *47* (48), 12729-12739.
65. Cheung-Lau, J. C.; Liu, D.; Pulsipher, K. W.; Liu, W.; Dmochowski, I. J., Engineering a well-ordered, functional protein-gold nanoparticle assembly. *Journal of Inorganic Biochemistry* **2014,** *130*, 59-68.
66. Kramer, R. M.; Li, C.; Carter, D. C.; Stone, M. O.; Naik, R. R., Engineered protein cages for nanomaterial synthesis. *Journal of the American Chemical Society* **2004,** *126* (41), 13282-13286.
67. Galvez, N.; Sanchez, P.; Dominguez-Vera, J. M., Preparation of Cu and CuFe Prussian Blue derivative nanoparticles using the apoferritin cavity as nanoreactor. *Dalton Transactions* **2005,** (15), 2492-2494.
68. Yamashita, I.; Hayashi, J.; Hara, M., Bio-template synthesis of uniform CdSe nanoparticles using cage-shaped protein, apoferritin. *Chemistry Letters* **2004,** *33* (9), 1158-1159.
69. Ueno, T.; Suzuki, M.; Goto, T.; Matsumoto, T.; Nagayama, K.; Watanabe, Y., Size-selective olefin hydrogenation by a Pd nanocluster provided in an apo-ferritin cage. *Angewandte Chemie-International Edition* **2004,** *43* (19), 2527-2530.
70. Zhang, L.; Laug, L.; Muenchgesang, W.; Pippel, E.; Goesele, U.; Brandsch, M.; Knez, M., Reducing Stress on Cells with Apoferritin-Encapsulated Platinum Nanoparticles. *Nano Letters* **2010,** *10* (1), 219-223.
71. Daniel, M. C.; Astruc, D., Gold Nanoparticles: Assembly, Supramolecular Chemistry, Quantum-Size-Related Properties, and Applications Toward Biology, Catalysis, and Nanotechnology. *Chemical Reviews* **2004,** *104* (1), 293-346.

72. Shenton, W.; Mann, S.; Colfen, H.; Bacher, A.; Fischer, M., Synthesis of nanophase iron oxide in lumazine synthase capsids. *Angewandte Chemie-International Edition* **2001,** *40* (2), 442-445.
73. Douglas, T.; Young, M., Host-guest encapsulation of materials by assembled virus protein cages. *Nature* **1998,** *393* (6681), 152-155.
74. Woersdoerfer, B.; Woycechowsky, K. J.; Hilvert, D., Directed Evolution of a Protein Container. *Science* **2011,** *331* (6017), 589-592.
75. Sana, B.; Johnson, E.; Sheah, K.; Poh, C. L.; Lim, S., Iron-based ferritin nanocore as a contrast agent. *Biointerphases* **2010,** *5* (3), FA48-FA52.
76. Schoonen, L.; van Hest, J. C. M., Functionalization of protein-based nanocages for drug delivery applications. *Nanoscale* **2014,** *6* (13), 7124-7141.
77. Ruder, W. C.; Lu, T.; Collins, J. J., Synthetic Biology Moving into the Clinic. *Science* **2011,** *333* (6047), 1248-1252.
78. Pierce, M. M.; Raman, C. S.; Nall, B. T., Isothermal titration calorimetry of protein-protein interactions. *Methods-a Companion to Methods in Enzymology* **1999,** *19* (2), 213-221.
79. Velazquez-Campoy, A.; Leavitt, S. A.; Freire, E., Characterization of protein-protein interactions by isothermal titration calorimetry. *Methods in molecular biology (Clifton, N.J.)* **2004,** *261*, 35-54.
80. Chiu, M. H.; Prenner, E. J., Differential scanning calorimetry: An invaluable tool for a detailed thermodynamic characterization of macromolecules and their interactions. *Journal of pharmacy & bioallied sciences* **2011,** *3* (1), 39-59.
81. Brandts, J. F.; Lin, L. N., Study of strong to ultratight protein interactions using differential scanning calorimetry. *Biochemistry* **1990,** *29* (29), 6927-6940.
82. Figeys, D.; McBroom, L. D.; Moran, M. F., Mass spectrometry for the study of protein-protein interactions. *Methods* **2001,** *24* (3), 230-239.
83. Gupta, P. K.; Rustgi, S., Mass spectrometry: An essential tool for genome and proteome analysis. *Indian Journal of Biotechnology* **2003,** *2* (1), 48-64.
84. Song, J.; Wang, Z.; Ewing, R. M., Integrated analysis of the Wnt responsive proteome in human cells reveals diverse and cell-type specific networks. *Molecular Biosystems* **2014,** *10* (1), 45-53.
85. Aebersold, R.; Mann, M., Mass spectrometry-based proteomics. *Nature* **2003,** *422* (6928), 198-207.
86. Homola, J.; Yee, S. S.; Gauglitz, G., Surface plasmon resonance sensors: review. *Sensors and Actuators B-Chemical* **1999,** *54* (1-2), 3-15.
87. Cui, X. Q.; Yang, F.; Sha, Y. F.; Yang, X. R., Real-time immunoassay of ferritin using surface plasmon resonance biosensor. *Talanta* **2003,** *60* (1), 53-61.
88. Santini, S.; Di Agostino, S.; Coppari, E.; Bizzarri, A. R.; Blandino, G.; Cannistraro, S., Interaction of mutant p53 with p73: A Surface Plasmon Resonance and Atomic Force Spectroscopy study. *Biochimica Et Biophysica Acta-General Subjects* **2014,** *1840* (6), 1958-1964.
89. Giepmans, B. N. G.; Adams, S. R.; Ellisman, M. H.; Tsien, R. Y., Review - The fluorescent toolbox for assessing protein location and function. *Science* **2006,** *312* (5771), 217-224.
90. Towbin, H.; Staehelin, T.; Gordon, J., Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets - procedure and some applications. *Proceedings of the National Academy of Sciences of the United States of America* **1979,** *76* (9), 4350-4354.
91. Salim, K.; Fenton, T.; Bacha, J.; Urien-Rodriguez, H.; Bonnert, T.; Skynner, H. A.; Watts, E., Oligomerization of G-protein-coupled receptors shown by selective co-immunoprecipitation. *Journal of Biological Chemistry* **2002,** *277* (18), 15482-15485.

92. Geva, G.; Sharan, R., Identification of protein complexes from co-immunoprecipitation data. *Bioinformatics* **2011,** *27* (1), 111-117.
93. Engvall, E.; Perlmann, P., Enzyme-linked immunosorbent assay (ELISA) quantitative assay of immunoglobulin-G. *Immunochemistry* **1971,** *8* (9), 871-&.
94. Lequin, R. M., Enzyme Immunoassay (EIA)/Enzyme-Linked Immunosorbent Assay (ELISA). *Clinical Chemistry* **2005,** *51* (12), 2415-2418.
95. Paltiel, A. D.; Weinstein, M. C.; Kimmel, A. D.; Seage, G. R.; Losina, E.; Zhang, H.; Freedberg, K. A.; Walensky, R. P., Expanded screening for HIV in the United States - An analysis of cost-effectiveness. *New England Journal of Medicine* **2005,** *352* (6), 586-595.
96. Jameson, D. M.; Seifried, S. E., Quantification of protein-protein interactions using fluorescence polarization. *Methods-a Companion to Methods in Enzymology* **1999,** *19* (2), 222-233.
97. Park, S.-H.; Raines, R. T., Fluorescence polarization assay to quantify protein-protein interactions. *Methods in molecular biology (Clifton, N.J.)* **2004,** *261*, 161-6.
98. Knight, S. M. G.; Umezawa, N.; Lee, H. S.; Gellman, S. H.; Kay, B. K., A fluorescence polarization assay for the identification of inhibitors of the p53-DM2 protein-protein interaction. *Analytical Biochemistry* **2002,** *300* (2), 230-236.
99. Richards, F. M., On the enzymic activity of subtilisin-modified ribonclease. *Proceedings of the National Academy of Sciences of the United States of America* **1958,** *44* (2), 162-166.
100. Shimomura, O.; Johnson, F. H.; Saiga, Y., Extraction, purification and properties of aequorin, a bioluminescent protein from luminous hydromedusan, aequorea. *Journal of Cellular and Comparative Physiology* **1962,** *59* (3), 223-&.
101. Prasher, D. C.; Eckenrode, V. K.; Ward, W. W.; Prendergast, F. G.; Cormier, M. J., Primary structure of the Aequorea-Victoria green-fluorescent protein. *Gene* **1992,** *111* (2), 229-233.
102. Chalfie, M., GFP: Lighting Up Life (Nobel Lecture). *Angewandte Chemie International Edition* **2009,** *48* (31), 5603-5611.
103. Shimomura, O., Discovery of Green Fluorescent Protein (GFP) (Nobel Lecture). *Angewandte Chemie International Edition* **2009,** *48* (31), 5590-5602.
104. Tsien, R. Y., Constructing and Exploiting the Fluorescent Protein Paintbox (Nobel Lecture). *Angewandte Chemie International Edition* **2009,** *48* (31), 5612-5626.
105. Heim, R.; Prasher, D. C.; Tsien, R. Y., Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proceedings of the National Academy of Sciences of the United States of America* **1994,** *91* (26), 12501-12504.
106. Cubitt, A. B.; Heim, R.; Adams, S. R.; Boyd, A. E.; Gross, L. A.; Tsien, R. Y., Understanding, improving and using green fluorescent proteins. *Trends in Biochemical Sciences* **1995,** *20* (11), 448-455.
107. Crameri, A.; Whitehorn, E. A.; Tate, E.; Stemmer, W. P. C., Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nature Biotechnology* **1996,** *14* (3), 315-319.
108. Cormack, B. P.; Valdivia, R. H.; Falkow, S., FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **1996,** *173* (1), 33-38.
109. Nagai, T.; Ibata, K.; Park, E. S.; Kubota, M.; Mikoshiba, K.; Miyawaki, A., A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nature Biotechnology* **2002,** *20* (1), 87-90.
110. Shaner, N. C.; Patterson, G. H.; Davidson, M. W., Advances in fluorescent protein technology. *Journal of Cell Science* **2007,** *120* (24), 4247-4260.
111. Rizzo, M. A.; Springer, G. H.; Granada, B.; Piston, D. W., An improved cyan fluorescent protein variant useful for FRET. *Nature Biotechnology* **2004,** *22* (4), 445-449.

112. Ghosh, I.; Hamilton, A. D.; Regan, L., Antiparallel leucine zipper-directed protein reassembly: Application to the green fluorescent protein. *Journal of the American Chemical Society* **2000,** *122* (23), 5658-5659.
113. Magliery, T. J.; Wilson, C. G. M.; Pan, W. L.; Mishler, D.; Ghosh, I.; Hamilton, A. D.; Regan, L., Detecting protein-protein interactions with a green fluorescent protein fragment reassembly trap: Scope and mechanism. *Journal of the American Chemical Society* **2005,** *127* (1), 146-157.
114. Rodriguez-Banqueri, A.; Kowalczyk, L.; Palacin, M.; Luis Vazquez-Ibar, J., Assessment of membrane protein expression and stability using a split green fluorescent protein reporter. *Analytical Biochemistry* **2012,** *423* (1), 7-14.
115. Barnard, E.; Timson, D. J., Split-EGFP screens for the detection and localisation of protein-protein interactions in living yeast cells. *Methods in molecular biology (Clifton, N.J.)* **2010,** *638*, 303-17.
116. Buck, C. B.; Cheng, N.; Thompson, C. D.; Lowy, D. R.; Steven, A. C.; Schiller, J. T.; Trus, B. L., Arrangement of L2 within the papillomavirus capsid. *Journal of Virology* **2008,** *82* (11), 5190-5197.
117. Dewet, J. R.; Wood, K. V.; Deluca, M.; Helinski, D. R.; Subramani, S., Firefly luciferase gene - structure and expression in mammalian-cells. *Molecular and Cellular Biology* **1987,** *7* (2), 725-737.
118. Jester, B. W.; Cox, K. J.; Gaj, A.; Shomin, C. D.; Porter, J. R.; Ghosh, I., A Coiled-Coil Enabled Split-Luciferase Three-Hybrid System: Applied Toward Profiling Inhibitors of Protein Kinases. *Journal of the American Chemical Society* **2010,** *132* (33), 11727-11735.
119. Stains, C. I.; Furman, J. L.; Porter, J. R.; Rajagopal, S.; Li, Y.; Wyatt, R. T.; Ghosh, I., A General Approach for Receptor and Antibody-Targeted Detection of Native Proteins Utilizing Split-Luciferase Reassembly. *Acs Chemical Biology* **2010,** *5* (10), 943-952.
120. Ozawa, T.; Kaihara, A.; Sato, M.; Tachihara, K.; Umezawa, Y., Split luciferase as an optical probe for detecting protein-protein interactions in mammalian cells based on protein splicing. *Analytical Chemistry* **2001,** *73* (11), 2516-2521.
121. Remy, I.; Michnick, S. W., A highly sensitive protein-protein interaction assay based on Gaussia luciferase. *Nature Methods* **2006,** *3* (12), 977-979.
122. Pietraszewska-Bogiel, A.; Gadella, T. W. J., FRET microscopy: from principle to routine technology in cell biology. *Journal of Microscopy* **2011,** *241* (2), 111-118.
123. Kramer, H. E. A.; Fischer, P., The Scientific Work of Theodor Forster: A Brief Sketch of his Life and Personality. *Chemphyschem* **2011,** *12* (3), 555-558.
124. Lohse, M. J.; Nuber, S.; Hoffmann, C., Fluorescence/Bioluminescence Resonance Energy Transfer Techniques to Study G-Protein-Coupled Receptor Activation and Signaling. *Pharmacological Reviews* **2012,** *64* (2), 299-336.
125. Sekar, R. B.; Periasamy, A., Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations. *Journal of Cell Biology* **2003,** *160* (5), 629-633.
126. Hoffmann, C.; Gaietta, G.; Bunemann, M.; Adams, S. R.; Oberdorff-Maass, S.; Behr, B.; Vilardaga, J. P.; Tsien, R. Y.; Eisman, M. H.; Lohse, M. J., A FlAsH-based FRET approach to determine G protein - coupled receptor activation in living cells. *Nature Methods* **2005,** *2* (3), 171-176.
127. Sorkin, A.; McClure, M.; Huang, F. T.; Carter, R., Interaction of EGF receptor and Grb2 in living cells visualized by fluorescence resonance energy transfer (FRET) microscopy. *Current Biology* **2000,** *10* (21), 1395-1398.
128. Siegel, R. M.; Frederiksen, J. K.; Zacharias, D. A.; Chan, F. K. M.; Johnson, M.; Lynch, D.; Tsien, R. Y.; Lenardo, M. J., Fas preassociation required for apoptosis signaling and dominant inhibition by pathogenic mutations. *Science* **2000,** *288* (5475), 2354-2357.

129. Fields, S.; Song, O.-k., A novel genetic system to detect protein–protein interactions. *Nature* **1989,** *340* (6230), 245-246.
130. Barabasi, A. L.; Oltvai, Z. N., Network biology: Understanding the cell's functional organization. *Nature Reviews Genetics* **2004,** *5* (2), 101-U15.
131. Rual, J. F.; Venkatesan, K.; Hao, T.; Hirozane-Kishikawa, T.; Dricot, A.; Li, N.; Berriz, G. F.; Gibbons, F. D.; Dreze, M.; Ayivi-Guedehoussou, N.; Klitgord, N.; Simon, C.; Boxem, M.; Milstein, S.; Rosenberg, J.; Goldberg, D. S.; Zhang, L. V.; Wong, S. L.; Franklin, G.; Li, S. M.; Albala, J. S.; Lim, J. H.; Fraughton, C.; Llamosas, E.; Cevik, S.; Bex, C.; Lamesch, P.; Sikorski, R. S.; Vandenhaute, J.; Zoghbi, H. Y.; Smolyar, A.; Bosak, S.; Sequerra, R.; Doucette-Stamm, L.; Cusick, M. E.; Hill, D. E.; Roth, F. P.; Vidal, M., Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **2005,** *437* (7062), 1173-1178.
132. Hu, J. C.; Kornacker, M. G.; Hochschild, A., Escherichia coli One- and Two-Hybrid Systems for the Analysis and Identification of Protein–Protein Interactions. *Methods* **2000,** *20* (1), 80-94.
133. Giot, L.; Bader, J. S.; Brouwer, C.; Chaudhuri, A.; Kuang, B.; Li, Y.; Hao, Y. L.; Ooi, C. E.; Godwin, B.; Vitols, E.; Vijayadamodar, G.; Pochart, P.; Machineni, H.; Welsh, M.; Kong, Y.; Zerhusen, B.; Malcolm, R.; Varrone, Z.; Collis, A.; Minto, M.; Burgess, S.; McDaniel, L.; Stimpson, E.; Spriggs, F.; Williams, J.; Neurath, K.; Ioime, N.; Agee, M.; Voss, E.; Furtak, K.; Renzulli, R.; Aanensen, N.; Carrolla, S.; Bickelhaupt, E.; Lazovatsky, Y.; DaSilva, A.; Zhong, J.; Stanyon, C. A.; Finley, R. L.; White, K. P.; Braverman, M.; Jarvie, T.; Gold, S.; Leach, M.; Knight, J.; Shimkets, R. A.; McKenna, M. P.; Chant, J.; Rothberg, J. M., A protein interaction map of Drosophila melanogaster. *Science* **2003,** *302* (5651), 1727-1736.
134. Skube, S. B.; Chaverri, J. M.; Goodson, H. V., Effect of GFP Tags on the Localization of EB1 and EB1 Fragments In Vivo. *Cytoskeleton* **2010,** *67* (1), 1-12.
135. Zhang, F. M.; Moniz, H. A.; Walcott, B.; Moremen, K. W.; Wang, L. C.; Linhardt, R. J., Probing the impact of GFP tagging on Robo1-heparin interaction. *Glycoconjugate Journal* **2014,** *31* (4), 299-307.
136. Griffin, B. A.; Adams, S. R.; Tsien, R. Y., Specific covalent labeling of recombinant protein molecules inside live cells. *Science* **1998,** *281* (5374), 269-272.
137. Hoffmann, C.; Gaietta, G.; Zuern, A.; Adams, S. R.; Terrillon, S.; Ellisman, M. H.; Tsien, R. Y.; Lohse, M. J., Fluorescent labeling of tetracysteine-tagged proteins in intact cells. *Nature Protocols* **2010,** *5* (10), 1666-1677.
138. Scheck, R. A.; Schepartz, A., Surveying Protein Structure and Function Using Bis-Arsenical Small Molecules. *Accounts of Chemical Research* **2011,** *44* (9), 654-665.
139. Stroffekova, K.; Proenza, C.; Beam, K. G., The protein-labeling reagent FLASH-EDT2 binds not only to CCXXCC motifs but also non-specifically to endogenous cysteine-rich proteins. *Pflugers Archiv-European Journal of Physiology* **2001,** *442* (6), 859-866.
140. Adams, S. R.; Campbell, R. E.; Gross, L. A.; Martin, B. R.; Walkup, G. K.; Yao, Y.; Llopis, J.; Tsien, R. Y., New biarsenical Ligands and tetracysteine motifs for protein labeling in vitro and in vivo: Synthesis and biological applications. *Journal of the American Chemical Society* **2002,** *124* (21), 6063-6076.
141. Van Engelenburg, S. B.; Nahreini, T.; Palmer, A. E., FACS-Based Selection of Tandem Tetracysteine Peptides with Improved ReAsH Brightness in Live Cells. *Chembiochem* **2010,** *11* (4), 489-493.
142. Martin, B. R.; Giepmans, B. N. G.; Adams, S. R.; Tsien, R. Y., Mammalian cell-based optimization of the biarsenical-binding tetracysteine motif for improved fluorescence and affinity. *Nature Biotechnology* **2005,** *23* (10), 1308-1314.
143. Soh, N., Selective chemical labeling of proteins with small fluorescent molecules based on metal-chelation methodology. *Sensors* **2008,** *8* (2), 1004-1024.

144. Ignatova, Z.; Gierasch, L. M., Monitoring protein stability and aggregation in vivo by real-time fluorescent labeling. *Proceedings of the National Academy of Sciences of the United States of America* **2004,** *101* (2), 523-528.
145. Taguchi, Y.; Shi, Z.-D.; Ruddy, B.; Dorward, D. W.; Greene, L.; Baron, G. S., Specific Biarsenical Labeling of Cell Surface Proteins Allows Fluorescent- and Biotin-tagging of Amyloid Precursor Protein and Prion Proteins. *Molecular Biology of the Cell* **2009,** *20* (1), 233-244.
146. Ray-Saha, S.; Schepartz, A., Visualizing Tyrosine Kinase Activity with Bipartite Tetracysteine Display. *Chembiochem* **2010,** *11* (15), 2089-2091.
147. Thorn, K. S.; Naber, N.; Matuska, M.; Vale, R. D.; Cooke, R., A novel method of affinity-purifying proteins using a bis-arsenical fluorescein. *Protein Science* **2000,** *9* (2), 213-217.
148. Zurn, A.; Klenk, C.; Zabel, U.; Reiner, S.; Lohse, M. J.; Hoffmann, C., Site-Specific, Orthogonal Labeling of Proteins in Intact Cells with Two Small Biarsenical Fluorophores. *Bioconjugate Chemistry* **2010,** *21* (5), 853-859.
149. Rutkowska, A.; Haering, C. H.; Schultz, C., A FlAsH-Based Cross-Linker to Study Protein Interactions in Living Cells. *Angewandte Chemie-International Edition* **2011,** *50* (52), 12655-12658.
150. Luedtke, N. W.; Dexter, R. J.; Fried, D. B.; Schepartz, A., Surveying polypeptide and protein domain conformation and association with FlAsH and ReAsH. *Nature Chemical Biology* **2007,** *3* (12), 779-784.
151. Doerr, A., Smaller can be better. *Nat Meth* **2008,** *5* (1), 6-7.
152. Goodman, J. L.; Fried, D. B.; Schepartz, A., Bipartite Tetracysteine Display Requires Site Flexibility for ReAsH Coordination. *Chembiochem* **2009,** *10* (10), 1644-1647.
153. Shaikh, T. A.; Bakus, R. C.; Parkin, S.; Atwood, D. A., Structural characteristics of 2-halo-1,3,2-dithiarsenic compounds and tris-(pentafluorophenylthio)-arsen. *Journal of Organometallic Chemistry* **2006,** *691* (9), 1825-1833.
154. Scheck, R. A.; Lowder, M. A.; Appelbaum, J. S.; Schepartz, A., Bipartite Tetracysteine Display Reveals Allosteric Control of Ligand-Specific EGFR Activation. *Acs Chemical Biology* **2012,** *7* (8), 1367-1376.
155. Avraham, R.; Yarden, Y., Feedback regulation of EGFR signalling: decision making by early and delayed loops. *Nature Reviews Molecular Cell Biology* **2011,** *12* (2), 104-117.
156. Lowder, M. A.; Appelbaum, J. S.; Hobert, E. M.; Schepartz, A., Visualizing protein partnerships in living cells and organisms. *Current Opinion in Chemical Biology* **2011,** *15* (6), 781-788.
157. Chen, I.; Ting, A. Y., Site-specific labeling of proteins with small molecules in live cells. *Current Opinion in Biotechnology* **2005,** *16* (1), 35-40.
158. Andresen, M.; Schmitz-Salue, R.; Jakobs, S., Short tetracysteine tags to beta-tubulin demonstrate the significance of small labels for live cell imaging. *Molecular Biology of the Cell* **2004,** *15* (12), 5616-5622.
159. Pomorski, A.; Otlewski, J.; Krezel, A., The High Zn-II Affinity of the Tetracysteine Tag Affects Its Fluorescent Labeling with Biarsenicals. *Chembiochem* **2010,** *11* (9), 1214-1218.
160. Enninga, J.; Mounier, J.; Sansonetti, P.; Van Nhieu, G. T., Secretion of type III effectors into host cells in real time. *Nature Methods* **2005,** *2* (12), 959-965.
161. Mota, L. J.; Holden, D. W., FlAsHlights on bacterial virulence proteins. *Nature Methods* **2005,** *2* (12), 898-899.
162. Romantsov, T.; Battle, A. R.; Hendel, J. L.; Martinac, B.; Wood, J. M., Protein Localization in Escherichia coli Cells: Comparison of the Cytoplasmic Membrane Proteins ProP, LacY, ProW, AqpZ, MscS, and MscL. *Journal of Bacteriology* **2010,** *192* (4), 912-924.

163. Romantsov, T.; Helbig, S.; Culham, D. E.; Gill, C.; Stalker, L.; Wood, J. M., Cardiolipin promotes polar localization of osmosensory transporter ProP in Escherichia coli. *Molecular Microbiology* **2007,** *64* (6), 1455-1465.
164. Xiong, Y.; Chen, B.; Shi, L.; Fredrickson, J. K.; Bigelow, D. J.; Squier, T. C., Targeted Protein Degradation of Outer Membrane Decaheme Cytochrome MtrC Metal Reductase in Shewanella oneidensis MR-1 Measured Using Biarsenical Probe CrAsH-EDT2. *Biochemistry* **2011,** *50* (45), 9738-9751.
165. List, B.; Barbas, C. F.; Lerner, R. A., Aldol sensors for the rapid generation of tunable fluorescence by antibody catalysis. *Proceedings of the National Academy of Sciences of the United States of America* **1998,** *95* (26), 15351-15355.
166. Walter, K. U.; Vamvaca, K.; Hilvert, D., An active enzyme constructed from a 9-amino acid alphabet. *Journal of Biological Chemistry* **2005,** *280* (45), 37742-37746.
167. Gibson, D. G.; Glass, J. I.; Lartigue, C.; Noskov, V. N.; Chuang, R.-Y.; Algire, M. A.; Benders, G. A.; Montague, M. G.; Ma, L.; Moodie, M. M.; Merryman, C.; Vashee, S.; Krishnakumar, R.; Assad-Garcia, N.; Andrews-Pfannkoch, C.; Denisova, E. A.; Young, L.; Qi, Z.-Q.; Segall-Shapiro, T. H.; Calvey, C. H.; Parmar, P. P.; Hutchison, C. A.; Smith, H. O.; Venter, J. C., Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science* **2010,** *329* (5987), 52-56.
168. Appella, D. H.; Christianson, L. A.; Karle, I. L.; Powell, D. R.; Gellman, S. H., β-Peptide Foldamers: Robust Helix Formation in a New Family of β-Amino Acid Oligomers. *Journal of the American Chemical Society* **1996,** *118* (51), 13071-13072.
169. Ho, S. P.; DeGrado, W. F., Design of a 4-helix bundle protein: synthesis of peptides which self-associate into a helical protein. *Journal of the American Chemical Society* **1987,** *109* (22), 6751-6758.
170. Regan, L.; Degrado, W. F., Characterization of a helical protein designed from 1st principles. *Science* **1988,** *241* (4868), 976-978.
171. Farid, T. A.; Kodali, G.; Solomon, L. A.; Lichtenstein, B. R.; Sheehan, M. M.; Fry, B. A.; Bialas, C.; Ennist, N. M.; Siedlecki, J. A.; Zhao, Z.; Stetz, M. A.; Valentine, K. G.; Anderson, J. L. R.; Wand, A. J.; Discher, B. M.; Moser, C. C.; Dutton, P. L., Elementary tetrahelical protein design for diverse oxidoreductase functions. *Nat Chem Biol* **2013,** *9* (12), 826-833.
172. Koder, R. L.; Anderson, J. L. R.; Solomon, L. A.; Reddy, K. S.; Moser, C. C.; Dutton, P. L., Design and engineering of an O2 transport protein. *Nature* **2009,** *458* (7236), 305-309.
173. Zaccai, N. R.; Chi, B.; Thomson, A. R.; Boyle, A. L.; Bartlett, G. J.; Bruning, M.; Linden, N.; Sessions, R. B.; Booth, P. J.; Brady, R. L.; Woolfson, D. N., A de novo peptide hexamer with a mutable channel. *Nature Chemical Biology* **2011,** *7* (12), 935-941.
174. Fletcher, J. M.; Harniman, R. L.; Barnes, F. R. H.; Boyle, A. L.; Collins, A.; Mantell, J.; Sharp, T. H.; Antognozzi, M.; Booth, P. J.; Linden, N.; Miles, M. J.; Sessions, R. B.; Verkade, P.; Woolfson, D. N., Self-Assembling Cages from Coiled-Coil Peptide Modules. *Science* **2013,** *340* (6132), 595-599.
175. Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Röthlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Barbas, C. F.; Hilvert, D.; Houk, K. N.; Stoddard, B. L.; Baker, D., De Novo Computational Design of Retro-Aldol Enzymes. *Science* **2008,** *319* (5868), 1387-1391.
176. Rothlisberger, D.; Khersonsky, O.; Wollacott, A. M.; Jiang, L.; DeChancie, J.; Betker, J.; Gallaher, J. L.; Althoff, E. A.; Zanghellini, A.; Dym, O.; Albeck, S.; Houk, K. N.; Tawfik, D. S.; Baker, D., Kemp elimination catalysts by computational enzyme design. *Nature* **2008,** *453* (7192), 190-195.
177. Fleishman, S. J.; Whitehead, T. A.; Ekiert, D. C.; Dreyfus, C.; Corn, J. E.; Strauch, E.-M.; Wilson, I. A.; Baker, D., Computational Design of Proteins Targeting the Conserved Stem Region of Influenza Hemagglutinin. *Science* **2011,** *332* (6031), 816-821.

178. King, N. P.; Sheffler, W.; Sawaya, M. R.; Vollmar, B. S.; Sumida, J. P.; André, I.; Gonen, T.; Yeates, T. O.; Baker, D., Computational Design of Self-Assembling Protein Nanomaterials with Atomic Level Accuracy. *Science* **2012,** *336* (6085), 1171-1174.
179. Leung, D. W., E. Chen, and D. V. Goeddel., A method for the random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique* **1989,** (1), 11-15.
180. Eckert, K. A.; Kunkel, T. A., DNA polymerase fidelity and the polymerase chain reaction. *PCR methods and applications* **1991,** *1* (1), 17-24.
181. Cadwell, R. C.; Joyce, G. F., Randomization of genes by PCR mutagenesis. *PCR methods and applications* **1992,** *2* (1), 28-33.
182. McCullum, E.; Williams, B. R.; Zhang, J.; Chaput, J., Random Mutagenesis by Error-Prone PCR. In *In Vitro Mutagenesis Protocols*, Braman, J., Ed. Humana Press: 2010; Vol. 634, pp 103-109.
183. Hogrefe, H., Fine-Tuning Enzyme Activity Through Saturation Mutagenesis. In *In Vitro Mutagenesis Protocols*, Braman, J., Ed. Humana Press: 2010; Vol. 634, pp 271-283.
184. Wang, D.; Zhao, C.; Cheng, R.; Sun, F. Z., Estimation of the mutation rate during error-prone polymerase chain reaction. *Journal of Computational Biology* **2000,** *7* (1-2), 143-158.
185. Chen, K. Q.; Arnold, F. H., Tuning the activity of an enzyme for unusual environments - sequential random mutagenesis of Subtilisin-E for catalysis in dimethylformamide. *Proceedings of the National Academy of Sciences of the United States of America* **1993,** *90* (12), 5618-5622.
186. You, L.; Arnold, F. H., Directed evolution of subtilisin E in Bacillus subtilis to enhance total activity in aqueous dimethylformamide. *Protein Engineering* **1996,** *9* (1), 77-83.
187. Moore, J. C.; Arnold, F. H., Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents. *Nature Biotechnology* **1996,** *14* (4), 458-467.
188. Farinas, Edgardo T.; Schwaneberg, U.; Glieder, A.; Arnold, Frances H., Directed Evolution of a Cytochrome P450 Monooxygenase for Alkane Oxidation. *Advanced Synthesis & Catalysis* **2001,** *343* (6-7), 601-606.
189. Glieder, A.; Farinas, E. T.; Arnold, F. H., Laboratory evolution of a soluble, self-sufficient, highly active alkane hydroxylase. *Nature Biotechnology* **2002,** *20* (11), 1135-1139.
190. Stemmer, W. P. C., Rapid evolution of a protein in-vitro by dna shuffling. *Nature* **1994,** *370* (6488), 389-391.
191. Erijman, A.; Dantes, A.; Bernheim, R.; Shifman, J. M.; Peleg, Y., Transfer-PCR (TPCR): A highway for DNA cloning and protein engineering. *Journal of Structural Biology* **2011,** *175* (2), 171-177.
192. Erijman, A.; Shifman, J. M.; Peleg, Y., A single-tube assembly of DNA using the transfer-PCR (TPCR) platform. *Methods in molecular biology (Clifton, N.J.)* **2014,** *1116*, 89-101.
193. Singh, M. I.; Jain, V., Tagging the Expressed Protein with 6 Histidines: Rapid Cloning of an Amplicon with Three Options. *Plos One* **2013,** *8* (5).
194. Wan, X. B.; Zhang, W.; Li, L.; Xie, Y. T.; Li, W.; Huang, N., A New Target for an Old Drug: Identifying Mitoxantrone as a Nanomolar Inhibitor of PIM1 Kinase via Kinome-Wide Selectivity Modeling. *Journal of Medicinal Chemistry* **2013,** *56* (6), 2619-2629.
195. Aizner, Y.; Sharabi, O.; Shirian, J.; Dakwar, G. R.; Risman, M.; Avraham, O.; Shifman, J., Mapping of the Binding Landscape for a Picomolar Protein-Protein Complex through Computation and Experiment. *Structure* **2014,** *22* (4), 636-645.
196. Dahiyat, B. I.; Mayo, S. L., De Novo Protein Design: Fully Automated Sequence Selection. *Science* **1997,** *278* (5335), 82-87.

197. Yosef, E.; Politi, R.; Choi, M. H.; Shifman, J. M., Computational Design of Calmodulin Mutants with up to 900-Fold Increase in Binding Specificity. *Journal of Molecular Biology* **2009,** *385* (5), 1470-1480.
198. Hayashi, N.; Welschof, M.; Zewe, M.; Braunagel, M.; Dubel, S.; Breitling, F.; Little, M., Simultaneous mutagenesis of antibody CDR regions by overlap extension and PCR. *Biotechniques* **1994,** *17* (2), 310-&.
199. Miyazaki, K.; Arnold, F. H., Exploring nonnatural evolutionary pathways by saturation mutagenesis: Rapid improvement of protein function. *Journal of Molecular Evolution* **1999,** *49* (6), 716-720.
200. Hibbert, E. G.; Baganz, F.; Hailes, H. C.; Ward, J. M.; Lye, G. J.; Woodley, J. M.; Dalby, P. A., Directed evolution of biocatalytic processes. *Biomolecular Engineering* **2005,** *22* (1-3), 11-19.
201. Hibbert, E. G.; Senussi, T.; Costelloe, S. J.; Lei, W.; Smith, M. E. B.; Ward, J. M.; Hailes, H. C.; Dalby, P. A., Directed evolution of transketolase activity on non-phosphorylated substrates. *Journal of Biotechnology* **2007,** *131* (4), 425-432.
202. Hibbert, E. G.; Senussi, T.; Smith, M. E. B.; Costelloe, S. J.; Ward, J. M.; Hailes, H. C.; Dalby, P. A., Directed evolution of transketolase substrate specificity towards an aliphatic aldehyde. *Journal of Biotechnology* **2008,** *134* (3-4), 240-245.
203. Smith, M. E. B.; Hibbert, E. G.; Jones, A. B.; Dalby, P. A.; Hailes, H. C., Enhancing and Reversing the Stereoselectivity of Escherichia coli Transketolase via Single-Point Mutations. *Advanced Synthesis & Catalysis* **2008,** *350* (16), 2631-2638.
204. Granier, T.; Gallois, B.; Dautant, A.; Destaintot, B. L.; Precigoux, G., Comparison of the structures of the cubic and tetragonal forms of horse-spleen apoferritin. *Acta Crystallographica Section D-Biological Crystallography* **1997,** *53*, 580-587.
205. de Val, N.; Declercq, J.-P.; Lim, C. K.; Crichton, R. R., Structural analysis of haemin demetallation by L-chain apoferritins. *Journal of Inorganic Biochemistry* **2012,** *112*, 77-84.
206. Stefanini, S.; Chiancone, E.; Arosio, P.; Finazziagro, A.; Antonini, E., Structural heterogeneity and subunit composition of horse ferritins. *Biochemistry* **1982,** *21* (10), 2293-2299.
207. Yariv, J.; Kalb, A. J.; Sperling, R.; Bauminger, E. R.; Cohen, S. G.; Ofer, S., The composition and the structure of bacterioferritin of Escherichia-coli. *Biochemical Journal* **1981,** *197* (1), 171-175.
208. Lebrun, N. E.; Andrews, S. C.; Guest, J. R.; Harrison, P. M.; Moore, G. R.; Thomson, A. J., Identification of the ferroxidase center of Escherichia-coli bacterioferritin. *Biochemical Journal* **1995,** *312*, 385-392.
209. Lawson, D. M.; Treffry, A.; Artymiuk, P. J.; Harrison, P. M.; Yewdall, S. J.; Luzzago, A.; Cesareni, G.; Levi, S.; Arosio, P., Idenitifcation of the ferroxidase active site in Ferritin. *Febs Letters* **1989,** *254* (1-2), 207-210.
210. Crow, A.; Lawson, T. L.; Lewin, A.; Moore, G. R.; Le Brun, N. E., Structural Basis for Iron Mineralization by Bacterioferritin. *Journal of the American Chemical Society* **2009,** *131* (19), 6808-6813.
211. Almiron, M.; Link, A. J.; Furlong, D.; Kolter, R., A novel DNA-binding protein with regulatory and protective roles in starved Escherichia-coli. *Genes & Development* **1992,** *6* (12B), 2646-2654.
212. Martinez, A.; Kolter, R., Protection of DNA during oxidative stress by the nonspecific DNA-binding protein Dps. *Journal of Bacteriology* **1997,** *179* (16), 5188-5194.
213. Grant, R. A.; Filman, D. J.; Finkel, S. E.; Kolter, R.; Hogle, J. M., The crystal structure of Dps, a ferritin homolog that binds and protects DNA. *Nature Structural Biology* **1998,** *5* (4), 294-303.

214. Chiancone, E.; Ceci, P., The multifaceted capacity of Dps proteins to combat bacterial stress conditions: Detoxification of iron and hydrogen peroxide and DNA binding. *Biochimica Et Biophysica Acta-General Subjects* **2010,** *1800* (8), 798-805.
215. Ueno, T.; Suzuki, M.; Goto, T.; Matsumoto, T.; Nagayama, K.; Watanabe, Y., Size-selective olefin hydrogenation by a Pd nanocluster provided in an apo-ferritin cage. *Angewandte Chemie-International Edition* **2004,** *43* (19), 2527-2530.
216. Yamashita, I.; Hayashi, J.; Hara, M., Bio-template synthesis of uniform CdSe nanoparticles using cage-shaped protein, apoferritin. *Chemistry Letters* **2004,** *33* (9), 1158-1159.
217. Butts, C. A.; Swift, J.; Kang, S. G.; Di Costanzo, L.; Christiansen, D. W.; Saven, J. G.; Dmochowski, I. J., Directing noble metal ion chemistry within a designed ferritin protein. *Biochemistry* **2008,** *47* (48), 12729-12739.
218. Cheung-Lau, J. C.; Liu, D.; Pulsipher, K. W.; Liu, W.; Dmochowski, I. J., Engineering a well-ordered, functional protein-gold nanoparticle assembly. *Journal of Inorganic Biochemistry* **2014,** *130*, 59-68.
219. Galvez, N.; Sanchez, P.; Dominguez-Vera, J. M., Preparation of Cu and CuFe Prussian Blue derivative nanoparticles using the apoferritin cavity as nanoreactor. *Dalton Transactions* **2005,** (15), 2492-2494.
220. Sana, B.; Johnson, E.; Sheah, K.; Poh, C. L.; Lim, S., Iron-based ferritin nanocore as a contrast agent. *Biointerphases* **2010,** *5* (3), FA48-FA52.
221. Li, K.; Zhang, Z.-P.; Luo, M.; Yu, X.; Han, Y.; Wei, H.-P.; Cui, Z.-Q.; Zhang, X.-E., Multifunctional ferritin cage nanostructures for fluorescence and MR imaging of tumor cells. *Nanoscale* **2012,** *4* (1), 188-193.
222. Mitchell, N.; Kalber, T. L.; Cooper, M. S.; Sunassee, K.; Chalker, S. L.; Shaw, K. P.; Ordidge, K. L.; Badar, A.; Janes, S. M.; Blower, P. J.; Lythgoe, M. F.; Hailes, H. C.; Tabor, A. B., Incorporation of paramagnetic, fluorescent and PET/SPECT contrast agents into liposomes for multimodal imaging. *Biomaterials* **2013,** *34* (4), 1179-1192.
223. Lin, X.; Xie, J.; Niu, G.; Zhang, F.; Gao, H.; Yang, M.; Quan, Q.; Aronova, M. A.; Zhang, G.; Lee, S.; Leaprnan, R.; Chen, X., Chimeric Ferritin Nanocages for Multiple Function Loading and Multimodal Imaging. *Nano Letters* **2011,** *11* (2), 814-819.
224. Uchida, M.; Flenniken, M. L.; Allen, M.; Willits, D. A.; Crowley, B. E.; Brumfield, S.; Willis, A. F.; Jackiw, L.; Jutila, M.; Young, M. J.; Douglas, T., Targeting of Cancer Cells with Ferrimagnetic Ferritin Cage Nanoparticles. *Journal of the American Chemical Society* **2006,** *128* (51), 16626-16633.
225. Kitagawa, T.; Kosuge, H.; Uchida, M.; Dua, M. M.; Iida, Y.; Dalman, R. L.; Douglas, T.; McConnell, M. V., RGD-Conjugated Human Ferritin Nanoparticles for Imaging Vascular Inflammation and Angiogenesis in Experimental Carotid and Aortic Disease. *Molecular Imaging and Biology* **2012,** *14* (3), 315-324.
226. Bryce, C. F. A.; Crichton, R. R., Catalytic activity of horse spleen apoferritin - preliminary kinetic studies and effect of chemical modification. *Biochemical Journal* **1973,** *133* (2), 301-&.
227. Kelly, S. M.; Jess, T. J.; Price, N. C., How to study proteins by circular dichroism. *Biochimica Et Biophysica Acta-Proteins and Proteomics* **2005,** *1751* (2), 119-139.
228. Greenfield, N. J., Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols* **2006,** *1* (6), 2876-2890.
229. Dautant, A.; Meyer, J. B.; Yariv, J.; Precigoux, G.; Sweet, R. M.; Kalb, A. J.; Frolow, F., Structure of a monoclinic crystal form of cytochrome b1 (bacterioferritin) from E-coli. Acta Crystallographica Section D-Biological Crystallography **1998,** 54, 16-24.
230. Grant, R. A.; Filman, D. J.; Finkel, S. E.; Kolter, R.; Hogle, J. M., The crystal structure of Dps, a ferritin homolog that binds and protects DNA. Nature Structural Biology **1998,** 5 (4), 294-303.

231. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF chimera - A visualization system for exploratory research and analysis. Journal of Computational Chemistry **2004,** 25 (13), 1605-1612.
232. Luzzago, A.; Cesareni, G., Isolation of point mutations that affect the folding of the H-chain of human ferritin in Escherichia-coli. *Embo Journal* **1989,** *8* (2), 569-576.
233. Stein, N., CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *Journal of Applied Crystallography* **2008,** *41*, 641-643.
234. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., Phaser crystallographic software. *Journal of Applied Crystallography* **2007,** *40*, 658-674.
235. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta Crystallographica Section D-Biological Crystallography* **2010,** *66*, 486-501.
236. Murshudov, G. N.; Vagin, A. A.; Dodson, E. J., Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D-Biological Crystallography* **1997,** *53*, 240-255.
237. Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D-Biological Crystallography* **2010,** *66*, 213-221.
238. Otwinowski, Z.; Minor, W., Processing of X-ray diffraction data collected in oscillation mode. *Macromolecular Crystallography, Pt A* **1997,** *276*, 307-326.
239. Painter, J.; Merritt, E. A., Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallographica Section D-Biological Crystallography* **2006,** *62*, 439-450.
240. Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W., NIH Image to ImageJ: 25 years of image analysis. Nature Methods 2012, 9 (7), 671-675.
241. Avraham, R.; Yarden, Y., Feedback regulation of EGFR signalling: decision making by early and delayed loops. Nature Reviews Molecular Cell Biology 2011, 12 (2), 104-117.
242. Inc, A. S., Discovery Studio Modeling Environment, Release 4.0, San Diego. 2013.
243. Enninga, J.; Mounier, J.; Sansonetti, P.; Van Nhieu, G. T., Secretion of type III effectors into host cells in real time. *Nature Methods* **2005,** *2* (12), 959-965.
244. Ignatova, Z.; Gierasch, L. M., Monitoring protein stability and aggregation in vivo by real-time fluorescent labeling. *Proceedings of the National Academy of Sciences of the United States of America* **2004,** *101* (2), 523-528.
245. Romantsov, T.; Helbig, S.; Culham, D. E.; Gill, C.; Stalker, L.; Wood, J. M., Cardiolipin promotes polar localization of osmosensory transporter ProP in Escherichia coli. *Molecular Microbiology* **2007,** *64* (6), 1455-1465.
246. Romantsov, T.; Battle, A. R.; Hendel, J. L.; Martinac, B.; Wood, J. M., Protein Localization in Escherichia coli Cells: Comparison of the Cytoplasmic Membrane Proteins ProP, LacY, ProW, AqpZ, MscS, and MscL. *Journal of Bacteriology* **2010,** *192* (4), 912-924.
247. Grant, R. A.; Filman, D. J.; Finkel, S. E.; Kolter, R.; Hogle, J. M., The crystal structure of Dps, a ferritin homolog that binds and protects DNA. *Nature Structural Biology* 1998, *5* (4), 294-303.
248. Bryce, C. F. A.; Crichton, R. R., Catalytic activity of horse spleen apoferritin - preliminary kinetic studies and effect of chemical modification. *Biochemical Journal* 1973, *133* (2), 301-&.

249. Dautant, A.; Meyer, J. B.; Yariv, J.; Precigoux, G.; Sweet, R. M.; Kalb, A. J.; Frolow, F., Structure of a monoclinic crystal form of cytochrome b1 (bacterioferritin) from E-coli. *Acta Crystallographica Section D-Biological Crystallography* 1998, *54*, 16-24.
250. Sobolev, V.; Eyal, E.; Gerzon, S.; Potapov, V.; Babor, M.; Prilusky, J.; Edelman, M., SPACE: a suite of tools for protein structure prediction and analysis based on complementarity and environment. *Nucleic Acids Research* 2005, *33*, W39-W43.
251. Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J., Basic Local Alignment Search Tool. *Journal of Molecular Biology* 1990, *215* (3), 403-410.